RANKED SET SAMPLING

DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

Master of Philosophy

IN

STATISTICS

BY

MOHD. FAIZAN

UNDER THE SUPERVISION OF

PROF. A. H. KHAN

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH
ALIGARH MUSLIM UNIVERSITY
ALIGARH (INDIA)

2006

DEDICATED

TO MY

GRAND FATHER

ACKNOWLEDGEMENT

I bow in reverence to the Almighty 'Allah' whose benign benediction gave me the required zeal for the completion of this work.

First and foremost I would like to express my sincere gratitude, reverence and utmost regards to my Supervisor, **Prof. A.H. Khan** for his illuminating, scholarly guidance and creative supervision right from the inception to the culmination of this project. He has been a source of strength, inspiration and confidence to me.

I am also indebted to the Chairman, **Prof. M.Z. Khan** and my teachers Prof. M. J. Ahsan, Prof. M.M. Khalid, Mr. H.M. Islam, Dr. R.U. Khan, Dr. H. Athar and other faculty members; it is indeed their honest efforts, which has inculcated in me, the ability to think.

I am getting short of words to express my gratitude for my respected Sir, **Dr. M. Yaqub** who always lend a helping hand whenever I was in need and provided me the computer facility, which helped me to complete this dissertation in time.

Thanks are also due to all non-teaching staff members of the department for their untiring help and cooperation.

My sincere thanks to my seniors and my research colleagues who have encouraged me throughout this work. I am particularly grateful to Mr. Quazzafi Rabbani, Mr. Zaki Anwar, Mr. Jahangir Khan, Mr. Ahmed Yusuf, Mr Naji, and Mr. Ziaul Haque.

I also express my special thanks to my senior Mr. Maroof A. Khan for helping me throughout this period by his ever-helping attitude.

My efforts could not have been successful without my friends Mr. Adil Ghaznavi, Mr. Asif Iqubal and Mr. Mohd. Zubair.

Above all my heart goes out in reverence to my dear parents for their tremendous patience, forbearance and affection.

Mohd. Faizan
Department of Statistics & Oper. Res.
Aligarh Muslim University, Aligarh

CONTENTS

PREFACE

Ranked set sampling is an alternative to simple random sampling that can offer large improvements in precision. The original notion of ranked set sampling was proposed by McIntyer in 1952. Ranked set sampling has attracted practical interest in application areas such as agriculture, forestry, ecological and environmental science, and medical studies etc.. We, therefore, in this Dissertation have tried to compile the available results.

The subject matter of the dissertation has been arranged in five chapters. Chapter I is of introductory nature and discusses the concept and method of sampling of ranked set sample. Some basic definitions and results needed in subsequent chapters are also discussed.

Chapter II deals with the estimation of parameters of some distributions using ranked set sampling.

Chapter III embodies reliability estimation of exponential failure law using ranked set sampling.

In Chapter IV, the theory is developed to test the hypothesis based on ranked set sampling.

Last chapter deals with the concept of multi-layers ranked set sampling and concomitants and utilizes the relationship to develop results on regression analysis.

A comprehensive bibliography is given in the end.

CHAPTER I

INTRODUCTION

1. Order statistics

If the random variable $X_1, X_2, \cdots, X_n$ are arranged in ascending order of magnitude as $X_{1:n} \le X_{2:n} \le X_{3:n} \le \cdots \le X_{i:n} \le \cdots \le X_{n:n}$, then $X_{i:n}$ or $X_{(i)}$ is called the i^{th} order statistics of a size n. The terms $X_{1:n} = \min\,(X_1, X_2, \cdots, X_n)$ and $X_{n:n} = \max\,(X_1, X_2, \cdots, X_n)$ are called smallest and largest order statistics respectively.

1.1 Distribution of order statistics

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a continuous probability density function *pdf* $f(x)$ and cumulative distribution function *cdf* $F(x)$. Then the *pdf* of $X_{r:n}$ $1 \le r \le n$, the r^{th} order statistic, is given by

$$f_{r:n}(x) = C_{r:n}[F(x)]^{r-1}[1 - F(x)]^{n-r} f(x), \; -\infty \le x \le \infty \tag{1.1}$$

where

$$C_{r:n} = \frac{n!}{(r-1)!(n-r)!} = [B(r, n-r+1)]^{-1} \tag{1.2}$$

and *cdf*

$$F_{r:n}(x) = P(X_{r:n} \le x) = \sum_{i=r}^{n} \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i} \tag{1.3}$$

$$= C_{r:n} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\, dt \tag{1.4}$$

where equation (1.4) is incomplete beta function. For X continuous, equation (1.1) can be obtained from equation (1.4) by differentiating w. r. t. x

In particular

$$F_{1:n}(x) = 1 - [1 - F(x)]^n, \tag{1.5}$$

$$F_{n:n}(x) = [F(x)]^n \tag{1.6}$$

The joint *pdf* $X_{r:n}$ and $X_{s:n}$ $(1 \le r \le s \le n)$ is given by

$$f_{r,s:n}(x,y) = C_{r,s:n}[F(x)]^{r-1}[F(y) - F(x)]^{s-r-1}[1 - F(y)]^{n-s} f(x) f(y)$$
$$-\infty < x < y < \infty \tag{1.7}$$

where

$$C_{r,s:n} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} = [B(r, s-r, n-s+1)]^{-1} \tag{1.8}$$

$$F_{r,s:n}(x,y) = P(X_{r:n} \le x, X_{s:n} \le y)$$
$$= \sum_{j=s}^{n} \sum_{i=r}^{n} \frac{n!}{i!(j-i)!(n-j)!} [F(x)]^i [F(y) - F(x)]^{j-i} [1 - F(y)]^{n-j} \tag{1.9}$$

2. Concomitant variable

Let $(X_i, Y_i), i = 1,2,\cdots,n$ be n pairs of independent random variables having a common bivariate distribution corresponding to (X,Y). When the X_i's are arranged in non-decreasing order then the Y-variate associated with the r^{th} order statistics, denoted by $Y_{[r:n]}$, is called the concomitant of the r^{th} order statistic (David, 1981).

The occurrence of concomitants in a variety of contexts independently promoted also another term induced order statistics. (Bhattacharya, 1974 and Sen, 1976). David *et al.* (1977) studied the small sample theory of distribution and expected value of rank of $Y_{[r:n]}$. The exact and asymptotic distribution theory of $Y_{[r:n]}$ and of its rank was studied by Yang (1977) when $(X_i, Y_i), i = 1,2,\cdots,n$ are from an arbitrary absolutely continuous bivariate population. As pointed out by Sen (1981), linear functions of concomitants may also be viewed as mixed rank statistics (Ghosh and Sen, 1971).

There are cases in practical problems where the variable of interest, X, is hard to measure and difficult to rank as well but a concomitant variable, Y, can be easily measured. Then the concomitant variable can be used for the ranking of the sampling units. The RSS scheme adapted in this situation as follows. At the first stage of RSS, the concomitant variable is measured on each unit of the simple random samples, and the units are ranked according to the numerical order of their values of the concomitant variable. Then the measured X values at the second stage are the induced order statistics by the order of Y values. Let $Y_{(r:n)}$ denote the r^{th} order statistic of the Y's and $X_{(r:n)}$ denote its corresponding X. Let $f_{X|Y_{(r:n)}}(x|y)$ denote the conditional density function of X given $Y_{(r:n)} = y$ and $g_{(r:n)}(y)$ the marginal density function of $Y_{(r:n)}$. Then we have

$$f_{(r:n)}(x) = \int f_{X|Y_{(r:n)}}(x|y)\, g_{(r:n)}(y)\, dy.$$

It is easy to see that

$$f(x)=\int\sum_{r=1}^{n}\frac{1}{n}f_{X|Y_{(r:n)}}(x|y)\, g_{(r:n)}(y)\, dy$$

$$=\frac{1}{n}\sum_{r=1}^{n}f_{(r:n)}(x)$$

3. Ranked set sampling

Suppose $X_1, X_2, \cdots, X_n$ is a random sample from $F(x)$ with mean μ and a finite variance σ^2. Then standard nonparametric estimator of μ is $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ with $\text{var}(\overline{X})=\frac{\sigma^2}{n}$. In contrast to SRS, RSS uses only one observation, namely, $X_{1:n}\equiv X_{(11)}$, the lowest observation, from this set, then $X_{2:n}\equiv X_{(22)}$, the second lowest from other independent set of n observations, and finally $X_{n:n}\equiv X_{(nn)}$, the largest observation from a last set of n observations. This process can be described in a Table 1.1 as follows.

Table 1.1. Display of n^2 observations in n sets of n each

$X_{(11)}$	$X_{(12)}$	$\cdots$	$X_{(1(n-1))}$	$X_{(1n)}$
$X_{(21)}$	$X_{(22)}$	$\cdots$	$X_{(2(n-1))}$	$X_{(2n)}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$X_{(n1)}$	$X_{(n2)}$	$\cdots$	$X_{(n(n-1))}$	$X_{(nn)}$

The important point to emphasize is that although RSS requires identification of as many as n^2 experimental or sampling units, only n of them, namely, $\{X_{(11)}, X_{(22)}, \cdots, X_{(nn)}\}$, are actually measured, thus making a comparison of this sampling strategy with SRS of the same size n.

It is obvious that the new sample $X_{(11)}, X_{(22)}, \cdots, X_{(nn)}$, known in the literature as a *Ranked Set Sampling* (RSS), are independent but not identically distributed. Moreover, marginally, $X_{(ii)}$ is distributed as $X_{i:n}$, the i^{th} order statistic in a sample of size n from $F(x)$.

McIntyre (1952), who introduced ranked set sampling, did not supply any mathematical theory to support his suggestion. Takahasi and Wakimoto (1968) independently arrived at the same results and supplied the necessary statistical theory. They showed that mean of a ranked set sample

$$\hat{\mu}_{rss} = \frac{1}{n}\sum_{i=1}^{n} X_{(ii)} \tag{3.1}$$

is an unbiased estimator of the population mean and is more efficient than the usual sample mean μ under simple random sampling, that is when both estimators are constructed on the basis of the same number n of actual measurements,

$$\operatorname{var}(\hat{\mu}_{rss}) < \operatorname{var}(\overline{X}) \ ! \tag{3.2}$$

A direct proof of this variance inequality follows from the well-known positively associated property of the order statistics (Tukey, 1958., Bickel, 1967). Dell (1969) and Dell and Clutter (1972) provide the following explicit expression for the variance of $\hat{\mu}_{rss}$

$$Var(\hat{\mu}_{rss}) = \frac{1}{n}\left(\sigma^2 - \frac{1}{n}\sum_{i=1}^{n}(\mu_{(i)} - \mu)^2\right). \tag{3.3}$$

where $\mu_{(i)}$ is the mean of $X_{i:n}$.

David and Levine (1972) considered the case where the ranking is done on the basis of a numerical covariate, instead of judgment. Under certain assumptions, including normality, they obtain a formula expressing relative precision (RP) in terms of the squared coefficient of correlation between the covariate and the variate of the interest. Stokes (1976, 1977) has explored this model further.

Stokes (1980) also proposed an estimator of the population variance based on a ranked set sample and gave a useful review of some theoretical aspects of RSS Stokes (1986). Muttlak and McDonald (1990a and 1990b) developed the RSS theory when the experimental units are selected with size-biased probability of selection. Yu and Lam (1997) proposed regression-type RSS estimator of the population mean.

Bhoj and Ahsanullah (1996) derived the BLUE-RSS for generalized geometric distribution and showed that it is more efficient than the BLUE-OS. Hossain and Muttlak (2000) applied this method to some other location-scale distributions. They also obtained the BLUE-RSS of the population mean and compared it with some other popular estimators based on SRS and RSS. Balakrishnan and Li (2005) derived the BLUE-ORSS of location and scale parameters for two special cases of generalized geometric distribution and compared them with BLUR-OS and BLUE-RSS given by Bhoj and Ahsanullah (1996).

3.1 Reported applications

(i). Yanagawa and Chen (1980) mention that the RSS technique is regularly employed at the Pastoral Research Laboratory, CSIRO at Armidale, NSW, Australia. A plate with four holes is randomly thrown on a field, the pasture in the four holes is ranked by eye, and a hole is selected for quantification of pasture.

(ii). Yanagawa and Chen (1980) also mention that the method has been used to estimate rice crops in Okinawa, Japan. They attribute this information to Mizuno (1974).

(iii). In order to carry out a vegetation study, which involves determination of the number of stems, the size of shrubs may be consider as concomitant variable for ranking the number of stems (the main variable of interest). Muttlak and McDonald (1990a, b) have used the method in their study and have combined RSS with size based sampling to obtain improved estimators for the mean number of stems.

(iv). Attempts have also been made to incorporate the ranked set sampling protocol into the area of nonparametric studies. Stokes and Sager (1988) have obtained an RSS estimator of the population distribution function $F(.)$. They have shown that an empirical distribution function based on RSS in an unbiased estimator of $F(.)$ and the estimator possesses smaller variance than the one based on SRS. Stokes and Sager (1988) have also shown how to use the Kolmogrov-Smirnov statistic to obtain a confidence band for $F(.)$ based on RSS.

4. Ordered ranked set sampling

Suppose $X_{RSS} = \{X_{(11)}, X_{(22)}, \cdots, X_{(nn)}\}$ is an RSS from a population with *pdf* $f(x)$ and cumulative density function $F(x)$. Now ordering X_{RSS} in an ascending order of magnitude, we get $X_{ORSS} = \{X_{1:n}^{ORSS} \leq X_{2:n}^{ORSS} \leq \cdots \leq X_{n:n}^{ORSS}\}$, which is referred to as an ordered ranked set sampling (ORSS). By noting that $X_{(ii)}(i=1,\cdots,n)$ in X_{RSS} are independent and nonidentically distributed (INID) as $F_{1:n}(x),\cdots,F_{n:n}(x)$, respectively, and then using the forms of density and joint density functions of order statistics from INID variables (see, for example, Vaughan and Venables, 1972 and Balakrishnan, 1988), the *pdf* of $X_{r:n}^{ORSS}$ and the joint *pdf* of $X_{r:n}^{ORSS}$ and $X_{s:n}^{ORSS}$ $(1 \le r \le s \le n)$ can be written as (Balakrishnan and Li, 2005)

$$f_{r:n}^{ORSS}(x) = \sum_{(i_1,\cdots,i_n)\in p} \sum_{j_1=i_1}^{n} \cdots \sum_{j_{r-1}=i_{r-1}}^{n} \sum_{j_{r+1}=0}^{i_{r+1}-1} \cdots \sum_{j_n=0}^{i_n-1} W_r^* f_{r':n^2}(x),$$

$$-\infty < x < \infty, 1 \le r \le n, \tag{4.1}$$

and

$$f_{r,s:n}^{ORSS}(x,y) = \sum_{(i_1,\cdots,i_n)\in p} \sum_{j_1=i_1}^{n} \cdots \sum_{j_{r-1}=i_{r-1}}^{n} \sum_{j_r=i_r}^{n} \sum_{j_{r+1}=i_{r+1}}^{n} \cdots \sum_{j_{s-1}=i_{s-1}}^{n} \sum_{j_s=0}^{i_s-1},$$

$$-\infty < x < y < \infty, 1 \le r \le s \le n$$

$$\times \sum_{j_{s+1}=0}^{i_{s+1}-1} \cdots \sum_{j_n=0}^{i_{n-1}} \sum_{k_{r+1}=0}^{j_{r+1}-1} \cdots \sum_{k_{s-1}=0}^{j_{s-1}-1} \sum_{i_{r+1}=0}^{k_{r+1}} \cdots \sum_{i_{s-1}=0}^{k_{s-1}} V_{r,s} * f_{\hat{r},\hat{s}:\hat{n}}(x,y)$$

$$. \tag{4.2}$$

where $\sum_{(i_1,\cdots,i_n)\in p}$ denotes the summation over all $n!$ permutation $(i_1,i_2,\cdots,i_n)$ of $(1,2,\cdots,n)$, and

$$W_r^* = \frac{1}{(r-1)!(n-r)!}\left[\prod_{k=1,k\neq r}^{n} \binom{n}{j_k}\right]\left[i_r \binom{n}{i_r}\right] \frac{(r'-1)!(n^2-r')!}{(n^2)!},$$

$$r' = i_r + \sum_{k=1,k\neq r}^{n} j_k ,$$

$$V^*_{r,s} = V_{i,j,k,l} \frac{(-1)^{\sum_{a=1}^{n} j_a - \sum_{a=1}^{s-1} i_a} .[(\hat{r}-1)!(\hat{n}-\hat{r}-1)!(\hat{n}-\hat{s})!]}{[(r-1)!(s-r-1)!(n-s)!]\hat{n}!},$$

$$V_{i,j,k,l} = \left\{ \prod_{a=1}^{r-1} \binom{n}{j_a} \binom{j_a - 1}{j_a - i_a} \right\} \left\{ i_r \binom{n}{i_r} \binom{n - i_r}{j_r - i_r} \right\}$$

$$\times \left\{ \prod_{a=r+1}^{s-1} \binom{n}{j_a} \binom{j_a - 1}{j_a - i_a} \binom{k_a}{l_a} \right\} \left\{ i_s \binom{n}{i_s} \binom{i_s - 1}{j_s} \right\}$$

$$\times \left\{ \prod_{a=s+1}^{n} \binom{n - i_a + j_a}{j_a} \binom{n}{i_a - j_a - 1} \right\},$$

$$\hat{r} = \sum_{a=1}^{s-1} j_a - \sum_{a=r+1}^{s-1} l_a + 1 - (s - r),$$

$$\hat{s} = \sum_{a=1}^{s-1} j_a + 1,$$

$$\hat{n} = (n+1)(n-s+1) + \sum_{a=1}^{n} j_a - \sum_{a=s}^{n} i_a .$$

From equations (4.1) and (4.2), the moments of ORSS can be derived for the distribution $F(x)$.

5. Best linear unbiased estimator

We could seek the best linear unbiased estimators (BLUEs) in the classes of linear combinations of the ranked set sample values. Sinha *et al.* (1996) examine the BLUEs of μ and σ for the normal distribution

$N(\mu, \sigma^2)$, and of the scale parameter in a single-parameter exponential distribution.

If X follows the location scale distribution $F\left(\frac{X-\mu}{\sigma}\right)$ then $Z=\left(\frac{X-\mu}{\sigma}\right)$ follows the distribution $F(z)$. The order statistic $Z_{(r:i)}$ is related to the order statistic $X_{(r:i)}$ by $Z_{(r:i)}=\left(\frac{X_{(r:i)}-\mu}{\sigma}\right)$.

Let $\alpha_r = E\{Z_{(r:i)}\}$ $v_r = Var\{Z_{(r:i)}\}$

We can express the $X_{(r:i)}$'s in the form of a linear regression model as follows

$$X_{(r:i)} = \mu + \sigma\alpha_r + \varepsilon_{ri}$$

where ε_{ri} are independent random variables with $E\varepsilon_{ri}=0$ and $Var\,\varepsilon_{ri}=\sigma^2 v_r$.

Let $\boldsymbol{\beta} = (\mu, \sigma)'$,

$$\mathbf{X} = (X_{(11)},\cdots,X_{(1n)},\cdots,X_{(n1)},\cdots,X_{(nn)})'$$

$$U=\begin{bmatrix}\mathbf{1}'\alpha_1\mathbf{1}\\ \vdots \\ \mathbf{1}'\alpha_n\mathbf{1}\end{bmatrix}$$

$$W=\begin{bmatrix}v_1 I & \\ & v_n I\end{bmatrix}$$

where $\mathbf{1}$ is an n dimensional vectors whose elements are all 1 and I is an $n\times n$ identity matrix. By the theory of linear regression analysis, the BLUE of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (U'W^{-1}U)^{-1}U'W^{-1}X.$$

The variance of $\hat{\boldsymbol{\beta}}$ is given by

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2 (U'W^{-1}U)^{-1}.$$

5.1 BLUE-ORSS for location-scale distributions

Suppose $X_{ORSS} = \{X_{1:n}^{ORSS} \le X_{2:n}^{ORSS} \le \cdots \le X_{n:n}^{ORSS}\}$ is an ORSS from a location-scale distribution with location parameter μ and scale parameter $\sigma > 0$. Let $Z_{r:n}^{ORSS} = (X_{r:n}^{ORSS} - \mu)/\sigma$ be the standardized ORSS. Define $\mu_{r:n}^{ORSS} = E(Z_{r:n}^{ORSS})$, and $\sigma_{r:n}^{ORSS} = Cov(Z_{r:n}^{ORSS}, Z_{s:n}^{ORSS})$, $1 \le r \le s \le n$. Then

$$E(X_{r:n}^{ORSS}) = \mu + \sigma\,\mu_{r:n}^{ORSS}$$

and

$$Cov(X_{r:n}^{ORSS}, X_{s:n}^{ORSS}) = \sigma^2\,\sigma_{r,s:n}^{ORSS}.$$

Following Lloyd (1952), we have the BLUE-ORSS of $\boldsymbol{\theta} = (\mu, \sigma)'$ to be

$$\boldsymbol{\theta}^* = (\mathbf{B}'\Sigma^{-1}\mathbf{B})^{-1}\mathbf{B}'\Sigma^{-1}X_{ORSS}, \tag{5.1}$$

and the variance-covariance matrix of BLUE-ORSS θ^* to be

$$Cov(\boldsymbol{\theta}^*) = \sigma^2(\mathbf{B}'\Sigma^{-1}\mathbf{B})^{-1}, \tag{5.2}$$

where

$$\mathbf{B} = [1\ \mu^{ORSS}],\quad \Sigma = [\sigma_{r,s:n}^{ORSS}]\quad 1 = [1,1,\cdots,1]' \text{ and}$$

$$\boldsymbol{\mu}^{ORSS} = [\mu_{1:n}^{ORSS} \le \mu_{2:n}^{ORSS} \le \cdots \le \mu_{n:n}^{ORSS}]'.$$

When the underlying distribution is symmetric around μ, (5.1) and (5.2) can be simplified and written as

$$\begin{cases} \mu^* = (1'\Sigma^{-1}1)^{-1}1'\Sigma^{-1}X_{ORSS}, \\ \sigma^* = ((\mu^{ORSS})'\Sigma^{-1}\mu^{ORSS})^{-1}(\mu^{ORSS})'\Sigma^{-1}X_{ORSS}, \end{cases} \tag{5.3}$$

$$\begin{cases} Var(\mu^*) = \sigma^2(1'\Sigma^{-1}1)^{-1}, \\ Var(\sigma^*) = \sigma^2((\mu^{ORSS})'\Sigma^{-1}\mu^{ORSS})^{-1}, \end{cases} \tag{5.4}$$

and $Cov(\mu^*,\sigma^*) = 0$ in this case, i.e., the BLUE-ORSS of μ and σ are uncorrelated.

6. Reliability theory

6.1 Reliability function

Reliability of a unit or a system is defined as the probability that it will perform satisfactory for at least a specified period of time without a major breakdown.

If T is failure time of a unit, then the reliability of the unit is defined as

$$\begin{aligned} R(t) &= P(T > t) \\ &= 1 - F(t) \end{aligned}$$

6.2 Failure rate function

The rate at which failures occur in a certain time interval $[t_1, t_2]$ is called the failure rate during that interval. It is defined as the probability that a failure per unit time occurs in the interval, given that a failure has not occurred prior to t_1 the beginning of the interval. Thus the failure rate is given by

$$\lambda(t) = \frac{\int_{t_1}^{t_2} f(t)\, dt}{(t_2 - t_1)\int_{t_1}^{\infty} f(t)\, dt}$$

$$\lambda(t) = \frac{\int_{t_1}^{\infty} f(t)\, dt - \int_{t_2}^{\infty} f(t)\, dt}{(t_2 - t_1)\int_{t_1}^{\infty} f(t)\, dt}$$

If, we substitute $t_1 = t$ and $t_2 = t + \Delta t$, we get

$$\lambda(t) = \frac{R(t) - R(t + \Delta t)}{\Delta t . R(t)}$$

6.3 Hazard rate function

The hazard rate is defined as the limit of the failure rate as the length of the interval, $[t_1, t_2]$ approaches zero. Thus, it is instantaneous failure rate.

$$h(t) = \lim_{\Delta t \to 0} \frac{R(t) - R(t + \Delta t)}{\Delta t\, R(t)}$$

$$= \frac{1}{R(t)}\left[-\frac{d}{dt} R(t)\right]$$

$$= \frac{d \ln R(t)}{dt} = \frac{f(t)}{R(t)}$$

$h(t)\, dt$ represent the probability that a device of age t will fail in the small interval of time t to $t + \Delta t$. Hazard rate indicates the change in the failure rate over the life span of the device.

6.4 Mean time to failure

The expected life, is defined as

$$E(T) = \int_0^{\infty} t\, f(t)\, dt$$

is the mean time to failure,

where $f(t)$ is the *pdf* of T, the life time of the item. As the lifetime of an item has to be non-negative, we define $f(t)$ for $T \geq 0$.

7. Some distributions

7.1 Uniform distribution

$$f(x) = \frac{1}{b-a}, \qquad a \leq x \leq b$$

$$F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \leq x < b, \\ 1, & b \leq x. \end{cases}$$

$$E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}, \quad k > 0 \text{ is an integer.}$$

7.2 Normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \mu, \sigma^2 > 0, \ -\infty < x < \infty$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

The central moments of odd order are all zero. The central moment of even order are as follows

$$E(X-\mu)^{2n} = [(2n-1)(2n-3)\cdots 3.1]\sigma^{2n}.$$

7.3 Exponential distribution

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \ x > 0$$

$$F(x) = 1 - e^{-x/\beta}, \ x > 0$$

$$E(X^n) = n!\,\beta^n$$

7.4 Weibull distribution

$$f(x) = \frac{\alpha}{\beta}\,x^{\alpha-1}\,e^{-(x^{\alpha})/\beta} \qquad x > 0$$

$$F(x) = 1 - e^{-(x^{\alpha})/\beta} \qquad x > 0$$

Moments of standard Weibull distribution are given by

$$E(X^n) = \beta^{n/\alpha}\,\Gamma(1 + \frac{n}{\alpha})$$

At $\alpha = 2$ it is known as Rayleigh distribution whereas at $\alpha = 1$ it is Exponential distribution.

7.5 Extreme-value distribution

$$f(x) = \frac{1}{\sigma}\exp\left[\frac{x-\theta}{\sigma} - \exp\left(\frac{x-\theta}{\sigma}\right)\right], \qquad \sigma > 0, -\infty < x < \infty$$

$$F(x) = 1 - \exp\left[-\exp\left(\frac{x-\theta}{\sigma}\right)\right]$$

7.6 Logistic distribution

$$f(x) = \frac{\alpha e^{x}}{(1+e^{x})^{\alpha+1}}, \qquad -\infty < x < \infty$$

$$F(x) = 1 - \frac{1}{(1+e^{x})^{\alpha}}$$

CHAPTER II

ESTIMATION OF PARAMETERS USING RANKED SET SAMPLING

1. Introduction

In this chapter we explore the concept of ranked set sampling (RSS) and exploit its full potential for estimation of parameters in some specific parametric models, such as normal, exponential, logistic, Weibull and extreme value distribution, and discuss the usefulness of RSS methods as well as their suitable modification for estimation of relevant parameters. Most of variations of RSS from addressing such issues as (i) What is the best linear unbiased estimator (BLUE) based on $X_{(11)},\cdots,X_{(nn)}$? (ii) What is the best selection of n order statistics one each from n sets of n observation each? (iii) Why not use $X_{(11)},X_{(21)},\cdots,X_{(n1)}$, all smallest, or $X_{(1n)},X_{(2n)},\cdots,X_{(nn)}$, all largest? (iv) Can we get away with a smaller (*i.e.*, a partial) RSS? As expected, the answers to the above questions depend on the particular model under study as well as the nature of the parameter being estimated.

In Section 2, we discuss estimation of normal mean as well as normal variance using an RSS. For the estimation of normal variance σ^2, we explore the possibility of obtaining an improved estimator based on $X_{(11)},X_{(22)},\cdots,X_{(nn)}$. Compared to the standard nonparametric estimator $\hat{\sigma}^2=\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i-\overline{X})^2$, specifically, we investigate if

$$\frac{1}{(n-1)}\sum_{i=1}^{n}(X_{(ii)} - \hat{\mu}_{rss})^2$$

can be used fruitfully. In Section 3, we consider the problem of estimation of an exponential mean. In Section 4, we discuss the estimation of the location and scale parameters of a logistic distribution. Finally, Section 5 addresses the problem of estimation of the location and scale parameters of Weibull and extreme- value distributions.

2. Estimation of mean and variance in normal $N(\mu,\sigma^2)$ population

In this section we consider the situation when the underlying population is normal, and discuss the use of RSS for estimation of its mean and variance.

2.1 Estimation of mean: BLUE

We first address the issue of how best to use the RSS, namely, $X_{(11)}, X_{(22)}, \cdots, X_{(nn)}$, for estimation of μ. Recall that $\mathrm{E}(X_{(ii)}) = \mu + v_i\sigma$ and $Var(\mathrm{X}_{(ii)}) = \gamma_i\sigma^2$, where v_i and γ_i are respectively the expected value and the variance of the i^{th} order statistics in a sample of size n from a standard normal population. Obviously, $\sum_{i=1}^{n} v_i = 0$, and by the symmetry of the normal distribution, $\sum_{i=1}^{n}\frac{v_i}{\gamma_i} = 0$. It then easily follows that the BLUE of μ is given by

$$\hat{\mu}_{blue} = \frac{\sum_{i=1}^{n}\frac{\mathrm{X}_{(ii)}}{\gamma_i}}{\sum_{i=1}^{n}\frac{1}{\gamma_i}} \qquad (2.1)$$

with minimum variance $= \dfrac{\sigma^2}{\sum_{i=1}^{n} \frac{1}{\gamma_i}}$, which is always smaller than $\dfrac{\sigma^2}{n^2}\sum_{i=1}^{n}\gamma_i = \text{var}(\hat{\mu}_{RSS})$. Thus $\hat{\mu}_{blue}$ offers first improvement over the standard RSS estimator $\hat{\mu}_{rss}$.

Incidentally, we can also derive the BLUE of μ based on a partial RSS, namely, $X_{(11)},\ldots X_{(ll)}$, for $l<n$. Starting with $\sum_{i=1}^{l} c_i X_{(ii)}$ and minimizing $\text{var}\left(\sum_{i=1}^{l} c_i X_{(ii)}\right) = \sum_{i=1}^{l} c_i^2 v_i$ subject to the unbiasedness conditions: $\sum_{i=1}^{l} c_i = 1, \sum_{i=1}^{l} c_i v_i = 0$ leads to the BLUE of μ as

$$\tilde{\mu}_{blue}(l) = \frac{\left(\sum_{i=1}^{l}\frac{v_i^2}{\gamma_i}\right)\left(\sum_{i=1}^{l}\frac{X_{(ii)}}{\gamma_i}\right) - \left(\sum_{i=1}^{l}\frac{v_i}{\gamma_i}\right)\left(\sum_{i=1}^{l}X_{(ii)}\frac{v_i}{\gamma_i}\right)}{\left(\sum_{i=1}^{l}\frac{v_i^2}{\gamma_i}\right)\left(\sum_{i=1}^{l}\frac{1}{\gamma_i}\right) - \left(\sum_{i=1}^{l}\frac{v_i}{\gamma_i}\right)^2} \qquad (2.2)$$

with

$$\text{var}(\tilde{\mu}_{blue}) = \frac{\sigma^2\left(\sum_{i=1}^{l}\frac{v_i^2}{\gamma_i}\right)}{\left(\sum_{i=1}^{l}\frac{v_i^2}{\gamma_i}\right)\left(\sum_{i=1}^{l}\frac{1}{\gamma_i}\right) - \left(\sum_{i=1}^{l}\frac{v_i}{\gamma_i}\right)^2} \qquad (2.3)$$

2.2 Estimation of mean: which order statistic?

We next address the issue of the right selection of order statistics in the context of RSS, given that we must select one from each set of n

observations, there are n such sets, and that the resultant estimator of μ is unbiased .The following variance inequality for order statistics of a normal distribution is useful for our subsequent discussion. Its validity for $n \leq 50$ can be seen from an inspection of the Tables in Tietjen *et al.* (1977), and asymptotic validity follows from David and Groeneveld (1982). Throughout this chapter, the sample medians defined in the usual way as $X_{median:n} = X_{k+1:n}$ for $n = 2k+1$

$X_{median:n} = \frac{1}{2}(X_{k:n} + X_{k+1:n})$ for $n = 2k$.

Lemma 2.1 var $(X_{median:n}) \leq$ var $(X_{r:n})$ for any r and n.

In view of the above result, we can recommend the use of the median from each set of n observations, and the mean of all such medians as an estimator of μ, namely,

$$\hat{\mu}\,(n:n) = \frac{1}{n}\,[X^{(1)}_{median:n} + \cdots + X^{(n)}_{median:n}] \tag{2.4}$$

where $X^{(i)}_{median:n}$ is the sample median from the i^{th} row of the Table 1.1. Obviously, $\hat{\mu}(n:n)$ is unbiased for μ and by Lemma 2.1, much better than the mean based on an ordinary RSS.

Theorem 2.1 $\text{var}(\overline{X}_n) < \text{var}(X_{median:n}) < 2\,\text{var}(\overline{X}_n)$

In fact, as the proof of the Theorem demonstrates, for $n = 2m$ (even), a slightly stronger variance inequality holds, namely

$$\text{var}\,(X_{m:2m}) < 2\,\text{var}\,(\overline{X}_{2m}) \tag{2.5}$$

In view of Theorem 2.1 and the above inequality, it follows that

$$\hat{\mu}_{(2,n)} = \frac{1}{2}\left[X_{(1\,(\frac{n+1}{2}))} + X_{(2\,(\frac{n+1}{2}))}\right] \quad \text{for } n \text{ odd} \tag{2.6}$$

which is the average of only two suitably selected ranked set observations for n odd, and

$$\hat{\mu}^{*}_{(2,n)} = \frac{1}{2}\left[X_{(1\,(\frac{n}{2}))} + X_{(2\,(\frac{n+2}{2}))}\right] \quad \text{for } n \text{ even} \tag{2.7}$$

which is also based on only two ranked set observations for n even, will be better than the simple average of n observations, whatever be n. This is very powerful and interesting result.

2.3 Estimation of mean: why not all smallest or all largest?

We now discuss some other variations of RSS. In many experimental studies, locating the smallest or the largest of a set of observations could be very easy compared to locating the median as suggested above or locating all order statistics as required in McIntyre's setup. In this part we need to assume that σ^2 is known (=1, without any loss of generality), and propose $\tilde{\mu}_{\min} = \frac{1}{n}\sum_{i=1}^{n} X_{(i1)} - \nu_1$, the bias corrected mean of n smallest observation as an estimator of μ. More efficiently, we suggest the use of

$$\tilde{\mu}_{\min}(m) = \frac{1}{m}\sum_{i=1}^{m} X_{(i1)} - \nu_1 \tag{2.8}$$

based on $m\,(<n)$ smallest means. Note that E $(\tilde{\mu}_{\min}(m))=\mu$ and var $(\tilde{\mu}_{\min}(m))=\frac{\nu_1}{m}$. Analogously, we could use the average of the maximum if that is more convenient.

2.4 Estimation of mean: concept of expansion

We next explore another interesting idea of expansion by raising the possibility of observing $l\ (> n)$ experimental units but measuring at most $m\ (< n)$ of them. The object is to efficiently use these m measurements to get an estimator of μ which is an unbiased and has less variance compared to the sample mean based on n measurements. As an example, for $n = 2$, to dominate $\overline{X}$ with variance $= \frac{\sigma^2}{2}$, we can record one more observation (i.e., 3 observation in all) but use only the median (i.e., only one measurement) to yield $\text{var}\,(\hat{\mu}_{median:3}) = 0.45\sigma^2 < \frac{\sigma^2}{2}$!! It turns out that this is common phenomenon, and can be done for every sample size n.

2.5 Estimation of mean: ordinary RSS

For estimation of σ^2 based on an RSS of size n, we first observe that a common sense estimator, namely $\frac{1}{(n-1)}\sum_{i=1}^{n}(X_{(ii)} - \hat{\mu}_{rss})^2$ is not unbiased for σ^2 because

$$E\left[\sum_{i=1}^{n}(X_{(ii)} - \hat{\mu}_{rss})^2\right] = \left(1 - \frac{1}{n}\right) E\left[\sum_{i=1}^{n} X_{(ii)}^2\right] - \left(\frac{1}{n}\right) \sum_{i \neq j=1}^{n} E(X_{(ii)})\, E(X_{(jj)})$$

$$= \left(1 - \frac{1}{n}\right) E\left[\sum_{i=1}^{n} X_{(1i)}^2\right] - \left(\frac{1}{n}\right) \sum_{i \neq j=1}^{n} E(X_{(1i)})\, E(X_{(1j)})$$

$$> \left(1 - \frac{1}{n}\right) E\left[\sum_{i=1}^{n} X_{(1i)}^2\right] - \left(\frac{1}{n}\right) \sum_{i \neq j=1}^{n} E(X_{(1i)})\, E(X_{(1j)})$$

$$= E\left[\sum_{i=1}^{n}(X_{(1i)} - \overline{X})^2\right] = (n-1)\sigma^2 \tag{2.9}$$

We therefore determine c_n such that $\tilde{\sigma}_n^2 = c_n\left[\sum_{i=1}^{n}(X_{(ii)} - \hat{\mu}_{rss})^2\right]$ is an unbiased for σ^2, and subsequently compute var $(\tilde{\sigma}_n^2)$ as $k_n\sigma^4$.

where

$$c_n = \frac{1}{2+(2/\pi)}$$

and

$$k_n = \frac{2+(3/\pi)-(1/\pi^2)}{[1+(1/\pi)]^2}$$

2.6 Estimations of variance: modification of RSS

As in the case of estimation of μ, here also we have explored the possibility of using the medians, the smallest, and the largest observations from the Table (1.1) in order to come up with an improved estimator of σ^2 compared to $\hat{\sigma}^2_{ordinary}$. Our computations for $n=2$ reveal that the variance of $\hat{\sigma}^2_{smallest} = \dfrac{(X_{(11)} - X_{(21)})^2}{2\gamma_1}$, the unbiased estimator of σ^2 based on the two smallest order statistics (Table, 1.1), is given by $\left[\dfrac{3}{2\gamma_1^2} + \dfrac{1}{2}\right]\sigma^4 = 3.728\sigma^4$. This is more than $2\sigma^4 = \mathrm{var}(\hat{\sigma}^2_{ordinary})$!

Similarly, for $n=3$, use of $\hat{\sigma}^2_{median} = \dfrac{\sum_{i=1}^{3}(X_{(i2)} - \overline{X}_{(.2)})^2}{2\gamma_2}$, where

$\frac{1}{3}\sum_{i=1}^{3} X_{(i2)}$, which is unbiased estimator of σ^2 based on three medians, results in $\mathrm{var}(\hat{\sigma}^2_{median}) = 1.012\sigma^4$, which is more than σ^4! Thus, the use of the above variations of RSS does not seem to work for estimation of σ^2.

3. Estimation of an exponential mean

In this section, we assume that $X_1 \cdots X_n$ is a random sample from $f(x|\theta) = \frac{1}{\theta}\exp(-\frac{x}{\theta}),\ x > 0, \theta > 0$ (Sinha *et al.*, 1996). Recall that

$$E(X_{i:n}) = \sum_{j=1}^{i} \frac{\theta}{(n-j+1)} = \theta c_{i:n}, \text{ and}$$

$$\mathrm{var}\,(X_{i:n}) = \sum_{j=1}^{i} \frac{\theta^2}{(n-j+1)^2} = \theta^2 d_{i:n}.$$

where

$$c_{i:n} = \sum_{j=1}^{i} \frac{1}{(n-j+1)}, \text{ and } d_{i:n} = \sum_{j=1}^{i} \frac{1}{(n-j+1)^2}$$

3.1 Ordinary RSS

The traditional unbiased estimator of θ based on a SRS of size n is given by $\hat{\theta}_1 = \overline{X}$ with $\mathrm{var}(\hat{\theta}_1) = \frac{\theta^2}{n}$. The McIntyre unbiased estimator of θ based on the usual RSS as described before is given by

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{(ii)} \tag{3.1}$$

with

$$\operatorname{var}(\hat{\theta}^2) = \theta^2 \frac{1}{n^2} \sum_{i=1}^{n} d_{i:n} . \tag{3.2}$$

Trivially, $\operatorname{var}(\hat{\theta}_2) < \operatorname{var}(\hat{\theta}_1)$, as expected.

We now discuss a few variations of RSS. Since, for every i, $Y_{i:n} = \dfrac{X_{(ii)}}{c_{i:n}}$ provides an unbiased estimator of θ with $\operatorname{var}(Y_{i:n}) = \theta^2 \dfrac{d_{i:n}}{c_{i:n}^2} = \theta^2\ a_{i:n}$, the BLUE of θ based on the RSS is readily seen to be

$$\hat{\theta}_3 = \frac{\sum_{i=1}^{n} \frac{Y_{i:n}}{a_{i:n}}}{\sum_{i=1}^{n} \frac{1}{a_{i:n}}} \tag{3.3}$$

with

$$\operatorname{var}(\hat{\theta}_3) = \frac{\theta^2}{\sum_{i=1}^{n} \frac{1}{a_{i:n}}} \tag{3.4}$$

Obviously, $\operatorname{var}(\hat{\theta}_3) < \operatorname{var}(\hat{\theta}_2)$, and thus $\hat{\theta}_3$ offers a uniform improvement over McIntyre's $\hat{\theta}_2$. Incidentally, if we have only a partial RSS, namely, $X_{(11)}, \cdots, X_{(ll)}$ for some $l < n$, the above argument shows that the BLUE of θ based on $\{X_{(11,)} \cdots, X_{(ll)}\}$ is given by

$$\hat{\theta}_3\ (l) = \frac{\sum_{i=1}^{l} \frac{Y_{i:n}}{a_{i:n}}}{\sum_{i=1}^{l} \frac{1}{a_{i:n}}} \tag{3.5}$$

with

$$\text{var}\ (\hat{\theta}_3(l)) = \frac{\theta^2}{\sum_{i=1}^{l} \frac{1}{a_{i:n}}} \tag{3.6}$$

3.3. Which order statistic?

We next address the issue of the right selection of order statistics, given that we must select exactly one from each set of n observations. It is clear from the preceding discussion that $Y_{r:n}$ is the best choice, where the index r makes $a_{r:n}$ the smallest $a_{1:n}, \cdots, a_{n:n}$. In the notation of Table 1.1, our proposed estimator of θ is given by

$$\hat{\theta}_4 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_{(ir)}}{c_{r:n}} \right) \tag{3.7}$$

with

$$\text{var}(\hat{\theta}_4) = \frac{1}{n} \theta^2 a_{r:n} \tag{3.8}$$

More efficiently, as in the case of estimation of a normal mean, we propose measuring only $m\,(< n)$ r^{th} order statistics, where r is defined as above, and the use of

$$\hat{\theta}_4\ (m) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{X_{(ir)}}{c_{r:n}} \right) \tag{3.9}$$

as an estimator of θ. Obviously,

$$\text{var}\ (\hat{\theta}_4(m)) = \frac{1}{m} \theta^2 a_{r:n} \tag{3.10}$$

We now state the following result regarding optimum selection of m. Its proof appears in Sinha *et al.* (1996).

Theorem 3.1 Let r be such that $a_{r:n}$ is the smallest among $a_{1:n}, \cdots, a_{n:n}$. Then $a_{r:n} < \frac{2}{n}$.

In view of the above theorem, it follows that, irrespective of the value of n, once the optimum selection of the order statistics in each row of the Table 1.1 is made, it is enough to repeat the sampling process only once.

3.4 Why not all smallest or all largest?

We now discuss another variation of RSS, namely, if it makes sense to measure only the smallest or only the largest in Table 1.1 rather than across the diagonal as in RSS. Since $a_{1:n} = 1$, it follows trivially that $\frac{1}{n}\sum_{i=1}^{n} X_{(i1)}$, the mean of the n smallest order statistics, behaves just like $\overline{X}$, and hence offers no improvement. On the other hand, since $a_{n:n} < 1$ (by Cauchy-Schwartz inequality), it is clear that the use of

$$\hat{\theta}_5 = \frac{1}{n}\sum_{i=1}^{n} X_{(in)}, \tag{3.11}$$

the mean of the n largest order statistics, always results in an improved estimator of θ.

4. Estimation of location and scale parameters of a logistic distribution

This section is based on Lam *et al.* (1995). Here we apply the concept of RSS and its suitable modifications for estimation of location and scale parameters of a logistic distribution.

We note that the *pdf* of a logistic distribution can be written as

$$f(x|\theta,\sigma)=\frac{1}{\sigma}\cdot\frac{\exp\left(-\frac{x-\theta}{\sigma}\right)}{\left[1+\exp\left(-\frac{x-\theta}{\sigma}\right)\right]^2}, \qquad -\infty<x,\theta<\infty,\sigma>0 \qquad (4.1)$$

where θ is the location parameter and σ is the scale parameter. First we give discussion of some standard estimators of θ and σ based on a SRS of size n, namely, $X_1,\cdots,X_n$.

It is clear that the conventional maximum likelihood estimators of θ and σ are extremely difficult to get in the context of (4.1), and their sample properties are completely unknown. However, based on the order statistics $X_{1:n},\cdots,X_{n:n}$, Lloyd's (1952) best linear estimators (BLUE's) are quite popular. Throughout this section, we have taken the BLUE's of θ and σ as our main standards for comparison against RSS estimators. Using

$$E(X_{r:n})=\theta+c_{r:n}\sigma, \mathrm{var}(X_{r:n})=d_{rr:n}\sigma^2, \mathrm{cov}(X_{r:n},x_{s:n})=d_{rs:n}\sigma^2 \qquad (4.2)$$

where $c_{r:n}$ and $d_{rr:n}$ are respectively the mean and variance of $X_{r:n}$ and $d_{rs:n}$ is the covariance between $X_{r:n}$ and $X_{s:n}$ from a standard logistic distribution with $\theta=0$ and $\sigma=1$, the BLUEs of θ and σ are given by Lloyd (1952) and Balakrishnan and Cohen (1991).

$$\hat{\theta}_{blue}=\frac{\mathbf{1}'(\Sigma^*)^{-1}\mathbf{X}^*}{\mathbf{1}'(\Sigma^*)^{-1}\mathbf{1}} \qquad (4.3)$$

$$\hat{\sigma}_{blue}=\frac{\boldsymbol{\alpha}'(\Sigma^*)^{-1}\mathbf{X}^*}{\boldsymbol{\alpha}'(\Sigma^*)^{-1}\boldsymbol{\alpha}} \qquad (4.4)$$

In the above, $\mathbf{X}^*=(X_{1:n},\cdots,X_{n:n})', \boldsymbol{\alpha}=(c_{1:n},\cdots,c_{n:n})'$ and $\Sigma^*=\mathrm{var}(\mathbf{X}^*)$. Moreover,

$$\operatorname{var}(\hat{\sigma}_{blue}) = \frac{\sigma^2}{\mathbf{1}'(\Sigma^*)^{-1}\mathbf{1}} \tag{4.5}$$

$$\operatorname{var}(\hat{\theta}_{blue}) = \frac{\sigma^2}{\boldsymbol{\alpha}'(\Sigma^*)^{-1}\boldsymbol{\alpha}} \tag{4.6}$$

The Fisher information matrix is given by

$$I_n(\theta,\sigma) = \begin{bmatrix} \dfrac{n}{3\sigma^2} & 0 \\ 0 & \dfrac{cn}{\sigma^2} \end{bmatrix} \tag{4.7}$$

where

$$c = 2\int_0^\infty x^2 \, \frac{e^x(1-e^x)^2}{(1+e^x)} \, dx = 2.34 \tag{4.8}$$

which provides the Rao-Cramer Lower Bound (RCLB) of the variance of any unbiased estimator $d_1\hat{\theta} + d_2\hat{\sigma}$ of $d_1\theta + d_2\sigma$ as

$$RCLB\ (d_1\,\hat{\theta} + d_2\,\hat{\sigma}) = \frac{d_1^2\ 3\sigma^2}{n} + \frac{d_2^2\ \sigma^2}{c\,n} \tag{4.9}$$

This gives

$$RCLB\ (\theta) = \frac{3}{n}\sigma^2 \tag{4.10}$$

$$RCLB\ (\sigma) = \frac{1}{cn}\sigma^2 \tag{4.11}$$

4.1 Estimation of θ: ordinary RSS

Since $E(X) = \theta$, following McIntyre's concept, our first estimator of θ can be taken as

$$\hat{\theta}_{rss} = \frac{1}{n}\sum_{r=1}^{n} X_{(rr)} \tag{4.12}$$

with

$$\operatorname{var}(\hat{\theta}_{rss}) = \frac{1}{n^2}\sigma^2 \sum_{r=1}^{n} d_{rr:n} \tag{4.13}$$

4.2 Estimation of θ: BLUE based on RSS

We now address the issue of how best to use the RSS, namely, $X_{(11)},\cdots,X_{(nn)}$, for the estimation of θ. Recall that $E(X_{(ii)}) = \theta + c_{i:n}\sigma$ and $\operatorname{var}(X_{(ii)}) = d_{ii:n}\sigma^2$. Starting with $\sum_{i=1}^{n} c_i X_{(ii)}$ and var $(\sum_{i=1}^{n} c_i X_{(ii)})$ subject to the unbiasedness conditions: $\sum_{i=1}^{n} c_i = 1$, $\sum_{i=1}^{n} c_i c_{i:n} = 0$, leads to the BLUE of θ as

$$\tilde{\theta}_{blue} = \frac{\left(\sum_{i=1}^{n} \frac{X_{(ii)}}{d_{ii:n}}\right)\left(\sum_{i=1}^{n} \frac{c_{i:n}^2}{d_{ii:n}}\right) - \left(\sum_{i=1}^{n} \frac{c_{i:n}}{d_{ii:n}}\right)\left(\sum_{i=1}^{n} \frac{c_{i:n} X_{(ii)}}{d_{ii:n}}\right)}{\left(\sum_{i=1}^{n} \frac{1}{d_{ii:n}}\right)\left(\sum_{i=1}^{n} \frac{c_{i:n}^2}{d_{ii:n}}\right) - \left(\sum_{i=1}^{n} \frac{c_{i:n}}{d_{ii:n}}\right)^2} \tag{4.14}$$

with

$$\operatorname{var}(\tilde{\theta}_{blue}) = \sigma^2 \frac{\left(\sum_{i=1}^{n} \frac{c_{i:n}^2}{d_{ii:n}}\right)}{\left(\sum_{i=1}^{n} \frac{1}{d_{ii:n}}\right)\left(\sum_{i=1}^{n} \frac{c_{i:n}^2}{d_{ii:n}}\right) - \left(\sum_{i=1}^{n} \frac{c_{i:n}}{d_{ii:n}}\right)^2} \tag{4.15}$$

Using the fact that, for a logistic distribution,

$$\sum_{i=1}^{n} c_{i:n} = 0, \sum_{i=1}^{n} \frac{c_{i:n}}{d_{ii:n}} = 0 \tag{4.16}$$

we get

$$\widetilde{\theta}_{blue} = \frac{\sum_{i=1}^{n} \frac{X_{(ii)}}{d_{ii:n}}}{\sum_{i=1}^{n} \frac{1}{d_{ii:n}}}, \ \mathrm{var}(\widetilde{\theta}_{blue}) = \frac{\sigma^2}{\sum_{i=1}^{n} \frac{1}{d_{ii:n}}} \tag{4.17}$$

Table 2.1 provides a comparison of var $(\widetilde{\theta}_{rss})$, var $(\widetilde{\theta}_{blue})$ and $RCLB\ (\theta)$ for $n = 3,4,5,\cdots,10$. Without loss of generality, we have taken $\sigma^2 = 1$.

Table 2.1: Comparison of variances for estimation of θ

n	var $(\widetilde{\theta}_{rss})$	var $(\widetilde{\theta}_{blue})$	$RCLB(\theta)$
3	0.5968	0.5695	1.0000
4	0.3711	0.3379	0.7500
5	0.2553	0.2227	0.6000
6	0.1872	0.1576	0.5000
7	0.1438	0.1171	0.4286
8	0.1142	0.0905	0.3750
9	0.0931	0.0720	0.3333
10	0.0776	0.0586	0.3000

It is clear from Table 2.1 that, a comparison of var $(\tilde{\theta}_{rss})$ with the RCLB reveals another striking feature that var $(\tilde{\theta}_{rss})$ is smaller than RCLB for all n, so that $\tilde{\theta}_{rss}$ performs better than any unbiased estimator of θ based on SRS for all n. This fact can be proved theoretically as the following result shows. Its proof appears in Lam *et al.* (1995).

Theorem 4.1 $\mathrm{var}(\tilde{\theta}_{rss}) < \dfrac{3\sigma^2}{n}$ (RCLB for θ) for any n.

As in the previous sections, we can also derive the BLUE of θ based on a partial RSS (PRSS), namely, $X_{(11)},\cdots,X_{(ll)}$, for $l < n$. Starting with $\sum_{i=1}^{l} c_i X_{(ii)}$ and minimizing var $(\sum_{i=1}^{l} c_i X_{(ii)})$ subject to the unbiasedness condition: $\sum_{i=1}^{l} c_i = 1$, $\sum_{i=1}^{l} c_i c_{i:n} = 0$, leads to the same form as (4.13) with n replaced by l as the upper limit in all the summation. The expression for the variance of the resultant estimator $\tilde{\theta}_{blue}(prss, l)$ is exactly as in (4.14) with the above change.

4.3 Estimation of θ: which order statistic?

Once again, we address the issue of the right selection of order statistics in the context of RSS. The following variance inequality for order statistics of a Logistic distribution is useful for our subsequent discussion. Its validity follows from David and Groeneveld (1982). Recall that a sample

median is defined as $X_{median:n} = X_{m+1:2m+1}$ when $n = 2m+1$, and as $\frac{1}{2}[X_{m:2m} + X_{m+1:2m}]$ when $n = 2m$.

Lemma 4.1 var $(X_{median:n}) \leq$ var $(X_{(r:n)})$ for any r and n.

In view of the above result, we can recommend the use of the sample median from each row of n observations in Table 1.1, and the mean of all such medians as an estimator of θ. Slightly more efficiently, we propose measuring only m from the first row of Table 1.1, where $m \leq n$, and use

$$\widetilde{\theta}_{median:n}(m) = \frac{1}{m}[X^{(1)}_{median:n} + \cdots + X^{(m)}_{median:n}] \tag{4.18}$$

as an estimator of θ. Here $X^{(i)}_{median:n}$ is the sample median from the i^{th} row of Table 1.1.

The following result, whose proof again appears in Lam *et al.* (1995), shows that it is enough to measure only two experimental units to achieve universal dominance over any unbiased estimator of θ based on a SRS, whatever be n. This result is similar to in the previous section.

Theorem 4.2 (i) $\frac{\text{var}(X_{m+1:2m+1})}{2} < \frac{3\sigma^2}{(2m+1)}$ (RCLB for θ) for $n = 2m+1$. (ii) $\frac{\text{var}(X_{m:2m})}{2} < \frac{3\sigma^2}{(2m)}$ (RCLB for θ) for $n = 2m$.

4.4 Estimation of σ: BLUE based on RSS

In this section the problem of estimation of the scale parameter σ in (4.1), the use of RSS and its suitable variations results in much improved estimators compared to the use of the BLUE $(\hat{\sigma}_{blue})$ under SRS.

To derive the BLUE of σ based on the entire McIntyre sample $X_{(11)},\cdots,X_{(nn)}$, we minimize the variance of $\sum_{i=1}^{n} c_i X_{(ii)}$ subject to the unbiasedness conditions, $\sum_{i=1}^{n} c_i = 0$, $\sum_{i=1}^{n} c_i\, c_{i:n} = 1$. This results in

$$\tilde{\sigma}_{blue} = \frac{\left(\sum_{i=1}^{n} \frac{X_{(ii)}\, c_{i:n}}{d_{ii:n}}\right)\left(\sum_{i=1}^{n} \frac{1}{d_{ii:n}}\right) - \left(\sum_{i=1}^{n} \frac{c_{i:n}}{d_{ii:n}}\right)\left(\sum_{i=1}^{n} \frac{X_{(ii)}}{d_{ii:n}}\right)}{\left(\sum_{i=1}^{n} \frac{1}{d_{ii:n}}\right)\left(\sum_{i=1}^{n} \frac{c_{i:n}^2}{d_{ii:n}}\right) - \left(\sum_{i=1}^{n} \frac{c_{i:n}}{d_{ii:n}}\right)^2} \tag{4.19}$$

with

$$\operatorname{var}(\tilde{\sigma}_{blue}) = \sigma^2 \frac{\left(\sum_{i=1}^{n} \frac{1}{d_{ii:n}}\right)}{\left(\sum_{i=1}^{n} \frac{1}{d_{ii:n}}\right)\left(\sum_{i=1}^{n} \frac{c_{i:n}^2}{d_{ii:n}}\right) - \left(\sum_{i=1}^{n} \frac{c_{i:n}}{d_{ii:n}}\right)^2} \tag{4.20}$$

Using (4.16), the above can be simplified as

$$\tilde{\sigma}_{blue} = \frac{\sum_{i=1}^{n} \frac{X_{(ii)}\, c_{i:n}}{d_{i:n}}}{\sum_{i=1}^{n} \frac{c_{i:n}^2}{d_{i:n}}} \tag{4.21}$$

with

$$\operatorname{var}(\tilde{\sigma}_{blue}) = \sigma^2 \frac{1}{\sum_{i=1}^{n} \frac{c_{i:n}^2}{d_{i:n}}} \tag{4.22}$$

As in the case of estimation of θ, here also we can use a partial RSS, namely, $X_{(11)},\cdots,X_{(ll)}$ for some $l<n$. Starting with $\sum_{i=1}^{l} c_i X_{(ii)}$ and minimizing var $(\sum_{i=1}^{l} c_i X_{(ii)})$ subject to the usual unbiasedness conditions: $\sum_{i=1}^{l} c_i = 0$, $\sum_{i=1}^{l} c_i \, c_{i:n} = 1$, we can readily verify that the BLUE of σ based on the partial RSS is of the same form as (4.18) with n replaced by l in all the summations, and the corresponding variance of $\tilde{\sigma}_{blue}(prss,l)$ is also of the same form as (4.19) with the above change.

5. Estimation of parameters in Weibull and extreme-value distributions

This Section is based on the estimation of the parameters in a two parameters Weibull and extreme-value distributions (Fei *et al.*, 1994). We note that the *pdf* of a two-parameter Weibull distribution can be written as

$$f(y|\theta,\beta) = \frac{\beta}{\theta}\left(\frac{y}{\theta}\right)^{\beta-1} \exp\left[-\left(\frac{y}{\theta}\right)^{\beta}\right], \; y \geq 0 \tag{5.1}$$

where $\theta > 0, \beta > 0$ are the scale and shape parameters respectively.

Let $X = \ln Y$ (the natural logarithm of Y), then X has a Type I asymptotic distribution of smallest (extreme) values given by

$$g(x|\mu,\sigma) = \frac{1}{\sigma}\exp\left[\frac{x-\mu}{\sigma}\right]\exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right], -\infty < x < \infty \tag{5.2}$$

where $\mu = \ln\theta$, $\sigma = \frac{1}{\beta}$ are the location and scale parameters respectively.

If $Y_1, Y_2, \cdots, Y_n$ be a simple random sample (SRS) of size n from (5.1), then $X_1, X_2, \cdots, X_n$ are the SRS of size n from (5.2). Let $X_{1:n}, X_{2:n}, \cdots, X_{n:n}$ be the order statistics, then $Z_{i:n} = \frac{(X_{i:n} - \mu)}{\sigma}$, $i = 1, 2, \cdots, n$ are the order statistics from SRS of size n from a standard extreme-value distribution.

We shall use the notations

$$E(Z_{i:n}) = \alpha_{i:n}, \qquad i = 1, 2, \cdots, n, \tag{5.3}$$

$$Cov(Z_{i:n}, Z_{j:n}) = d_{ij}, \qquad i, j = 1, 2, \cdots, n \tag{5.4}$$

$$V = (d_{ij})_{n \times n}, \qquad V^{-1} = (d^{ij})_{n \times n}. \tag{5.5}$$

It is well known that the unique minimum variance linear unbiased estimators (UMVLUE) of μ and σ based on $X_1, \cdots, X_n$ (*i.e.*, $Y_1, \cdots, Y_n$) are given by

$$\hat{\mu} = \sum_{i=1}^{n} a_{i:n} X_{i:n} \tag{5.6}$$

$$\hat{\sigma} = \sum_{i=1}^{n} c_{i:n} X_{i:n} \tag{5.7}$$

The variance of the above estimator $\hat{\mu}$ and $\hat{\sigma}$ are as follows:

$$\operatorname{var}(\hat{\mu}) = A_n \sigma^2 \tag{5.8}$$

$$\operatorname{var}(\hat{\sigma}) = E_n \sigma^2 \tag{5.9}$$

and the covariance of $\hat{\mu}$ and $\hat{\sigma}$ can be written as

$$Cov(\hat{\mu}, \hat{\sigma}) = B_n \sigma^2 \tag{5.10}$$

where

$$A_n = \frac{1}{\Delta}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{i:n}\alpha_{j:n}d^{ij} \tag{5.11}$$

$$B_n = -\frac{1}{\Delta}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{i:n}d^{ij} \tag{5.12}$$

$$E_n = \frac{1}{\Delta}\sum_{i=1}^{n}\sum_{j=1}^{n}d^{ij} \tag{5.13}$$

Here

$$\Delta = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}d^{ij}\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{i:n}\alpha_{j:n}d^{ij}\right) - \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{i:n}d^{ij}\right)^2. \tag{5.14}$$

The coefficient $a_{i:n}$ and $c_{i:n}$ in the formulas (5.6) and (5.7) may be obtained as

$$a_{i:n} = A_n\sum_{j=1}^{n}d^{ij} + B_n\sum_{j=1}^{n}\alpha_{j:n}d^{ij} \tag{5.15}$$

$$c_{i:n} = B_n\sum_{j=1}^{n}d^{ij} + E_n\sum_{j=1}^{n}\alpha_{j:n}d^{ij} \tag{5.16}$$

We refer to Balakrishnan *et al.* (1992) for values of $\alpha_{i:n}'s, d_{ii}'s$ and $d_{ij}'s$, from which the above expressions can be evaluated. The coefficients $a_{i:n}$ and $c_{i:n}$ can also be obtained from Table 5.3 in Mann *et al.* (1974).

5.1 Estimation of μ and σ based on RSS

In this section we first discuss the problem of estimation of the parameters in (5.2) using a RSS. Clearly $X_{(11)}, X_{(22)}, \cdots, X_{(nn)}$ are independent, $X_{(ii)}$ is distributed as $X_{i:n}$, the i^{th} order statistic in a sample if size n from (5.2). Then, we have

$$E(X_{(ii)}) = \mu + \sigma E(Z_{i:n}) = \mu + \alpha_{i:n}\sigma, \quad i = 1,2,\cdots,n \tag{5.17}$$

$$Var(X_{(ii)}) = \sigma^2 Var(Z_{i:n}) = \sigma^2 d_{ii}, \quad i = 1,2,\cdots,n \tag{5.18}$$

$$Cov(X_{(ii)}, X_{(jj)}) = 0, \quad i \neq j, i, j = 1,2,\cdots,n. \tag{5.19}$$

We use the notations

$$\mathbf{X} = (X_{(11)}, X_{(22)}, \cdots, X_{(nn)})', \tag{5.20}$$

$$\Sigma = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix}, \tag{5.21}$$

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{d_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{d_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{d_{nn}} \end{pmatrix}, \tag{5.22}$$

$$E(\mathrm{X}) = \begin{pmatrix} \mu + \alpha_{1:n}\sigma \\ \mu + \alpha_{2:n}\sigma \\ \vdots \\ \mu + \alpha_{n:n}\sigma \end{pmatrix} = M\,\Theta \tag{5.23}$$

where $M = \begin{pmatrix} 1 & \alpha_{1:n} \\ 1 & \alpha_{2:n} \\ \vdots & \vdots \\ 1 & \alpha_{n:n} \end{pmatrix}$ and $\Theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$.

Using the generalized Gauss-Markov theorem, we can obtain unique minimum-variance linear estimators of μ and σ based on the RSS $(X_{(11)}, X_{(22)}, \cdots, X_{(nn)})$ as

$$\Theta^* = \begin{pmatrix} \mu^* \\ \sigma^* \end{pmatrix} = (M'\Sigma^{-1}M)^{-1} M'\Sigma^{-1} X \tag{5.24}$$

The variance-covariance matrix of the estimators μ^* and σ^* is given by

$$Var(\Theta^*) = \text{var}\begin{pmatrix} \mu^* \\ \sigma^* \end{pmatrix} = (M'\Sigma^{-1}M)^{-1}\sigma^2 \tag{5.25}$$

Furthermore, we may obtain the explicit formulas of the estimators μ^* and σ^* as

$$\mu^* = \frac{\left(\sum_{i=1}^{n} \frac{X_{(ii)}}{d_{ii}}\right)\left(\sum_{i=1}^{n} \frac{\alpha_{i:n}^2}{d_{ii}}\right) - \left(\sum_{i=1}^{n} \frac{\alpha_{i:n}}{d_{ii}}\right)\left(\sum_{i=1}^{n} \frac{\alpha_{i:n} X_{(ii)}}{d_{ii}}\right)}{\left(\sum_{i=1}^{n} \frac{1}{d_{ii}}\right)\left(\sum_{i=1}^{n} \frac{\alpha_{i:n}^2}{d_{ii}}\right) - \left(\sum_{i=1}^{n} \frac{\alpha_{i:n}}{d_{ii}}\right)^2} \tag{5.26}$$

and

$$\sigma^* = \frac{\left(\sum_{i=1}^{n} \frac{1}{d_{ii}}\right)\left(\sum_{i=1}^{n} \frac{\alpha_{i:n} X_{(ii)}}{d_{ii}}\right) - \left(\sum_{i=1}^{n} \frac{\alpha_{i:n}}{d_{ii}}\right)\left(\sum_{i=1}^{n} \frac{X_{(ii)}}{d_{ii}}\right)}{\left(\sum_{i=1}^{n} \frac{1}{d_{ii}}\right)\left(\sum_{i=1}^{n} \frac{\alpha_{i:n}^2}{d_{ii}}\right) - \left(\sum_{i=1}^{n} \frac{\alpha_{i:n}}{d_{ii}}\right)^2} \tag{5.27}$$

The variance of the estimator μ^* and σ^* are

$$Var(\mu^*) = \sigma^2 \frac{\left(\sum_{i=1}^{n} \frac{\alpha_{i:n}^2}{d_{ii}}\right)}{\left(\sum_{i=1}^{n} \frac{1}{d_{ii}}\right)\left(\sum_{i=1}^{n} \frac{\alpha_{i:n}^2}{d_{ii}}\right) - \left(\sum_{i=1}^{n} \frac{\alpha_{i:n}}{d_{ii}}\right)^2} \tag{5.28}$$

$$Var(\sigma^*) = \sigma^2 \frac{\left(\sum_{i=1}^{n} \frac{1}{d_{ii}}\right)}{\left(\sum_{i=1}^{n} \frac{1}{d_{ii}}\right)\left(\sum_{i=1}^{n} \frac{\alpha_{i:n}^2}{d_{ii}}\right) - \left(\sum_{i=1}^{n} \frac{\alpha_{i:n}}{d_{ii}}\right)^2} \tag{5.29}$$

We now discuss the use of PRSS in the context of estimation problems. Let $X_{(11)}, X_{(22)}, \cdots, X_{(ll)}$ $(l < n)$ be a partial RSS. It is easy to verify that the BLUE's of μ and σ based on the PRSS (denoted as $\tilde{\mu}_l$ and $\tilde{\sigma}_l$) are given exactly by (5.26) and (5.27) except that in all the summations, n is replaced by l. Morever, the corresponding variance of these BLUE's are given by (5.28) and (5.29) respectively with the same changes.

5.2 Relevance of smallest order statistic

In this section we study the possibility of using only the minimum order statistics and examine the performance of the of $(X_{(11)}, X_{(21)}, \cdots, X_{(l1)})$ for estimation of μ and σ for various choices of $l = 1, 2, \cdots, n$ in an attempt to find the minimum value of l for which dominance over $\hat{\mu}$ and $\hat{\sigma}$ holds.

It is easy to verify that $(X_{(11)}, X_{(21)}, \cdots, X_{(l1)})$ are *iid* with a common extreme-value distribution whose *pdf* is given by

$$g_{1,n}(x|u,\sigma) = \frac{1}{\sigma}\exp\left[\frac{x-u}{\sigma}\right]\exp\left[-\exp\left(\frac{x-u}{\sigma}\right)\right], -\infty < x < \infty \tag{5.30}$$

where the new location parameter is $u = \mu - \sigma \ln(n)$. To discuss estimation of μ and σ based on $(X_{(11)}, X_{(21)}, \cdots, X_{(l1)})$, let $X_{(1l)} < X_{(2l)} \cdots X_{(ll)}$ be their order statistics. Then the UMVLUE of u, σ may be obtained from (5.6), (5.7) as

$$\hat{u}_{1:l} = \sum_{i=1}^{l} a_{i:l} X_{(il)} \tag{5.31}$$

$$\hat{\sigma}_{1:l} = \sum_{i=1}^{l} c_{i:l} X_{(il)} \tag{5.32}$$

which results in the UMVLUE of μ as

$$\hat{\mu}_{1:l} = \hat{u}_{1:l} + \hat{\sigma}_{1:l} \ln(n) \tag{5.33}$$

By setting $n = l$ in (5.6) to (5.14), we immediately obtain the variance and co-variances of the estimators $\hat{u}_{1:l}$ and $\hat{\sigma}_{1:l}$ as

$$\mathrm{var}(\hat{u}_{1:l}) = A_l \, \sigma^2 \tag{5.34}$$

$$\mathrm{var}(\hat{\sigma}_{1:l}) = E_l \, \sigma^2 \tag{5.35}$$

$$Cov(\hat{u}_{1:l}, \hat{\sigma}_{1:l}) = B_l \, \sigma^2 \tag{5.36}$$

Then the variance of $\hat{\mu}_{1:l}$ and covariance of $\hat{\mu}_{1:l}$ and $\hat{\sigma}_{1:l}$ can be derived as

$$Var(\hat{\mu}_{1:l}) = [A_l + 2B_l \ln(n) + E_l (\ln(n))^2] \, \sigma^2 \tag{5.37}$$

$$\mathrm{Cov}(\hat{\mu}_{1:1}, \hat{\sigma}_{1:1}) = [\mathrm{B}_1 + E_l \ln(n)] \, \sigma^2 . \tag{5.38}$$

Using (5.11) to (5.14), we may obtain A_l, B_l, E_l and comparisons of $Var(\hat{\mu})$ with $Var(\hat{\mu}_{1:l})$ and $Var(\hat{\sigma})$ with $Var(\hat{\sigma}_{1:l})$ for various values of l and n. There is no improvement in estimation of μ and σ even in the best possible case of $l = n$.

CHAPTER III

RELIABILITY ESTIMATION BASED ON RANKED SET SAMPLING

1. Introduction

Consider a component whose lifetime is exponentially distributed random variable X with unknown mean $\theta > 0$. We wish to estimate the reliability of this component at an arbitrary but fixed point in time $t > 0$, namely $p = e^{-t/\theta}$. Let $X_1, \cdots, X_n$ be a random sample of size n representing the lifetime of n components. The minimum variance unbiased estimator of p is $\left[\left(1 - \frac{t}{T}\right)_+\right]^{n-1}$, where $T = \sum_{i=1}^{n} X_i$ and $(a)_+ = \max\,(a,\, 0)$. A more intuitive unbiased estimator for p in this case whose mathematical properties are more tractable is the empirical survival distribution function evaluated at t, namely $\frac{1}{n}\left(\sum_{i=1}^{n} I_{[X_i > t]}\right)$, where I_A is the indicator function of the set A. This unbiased estimator is simply the proportion of the n components that survive beyond time t (recall that $p = E(I_{[X > t]})$). Consider in this case the class of estimators of p which are of the form $\sum_{i=1}^{n} c_i I_{[X_i > \alpha t]}$, where $c_i \geq 0$, $i = 1, \cdots, n$, $\sum_{i=1}^{n} c_i = 1$ and $\alpha > 0$. The following facts about this class are straightforward: (i) An estimator in this class is unbiased if $\alpha = 1$, (ii) Among the infinite subclass of unbiased estimators, the empirical survival distribution function (evaluated at t) has the minimum variance, namely $\frac{p(1-p)}{n}$, which clearly decreases in n.

Let us now consider the task of estimating p based on RSS. It is clear that $\frac{1}{n}\left(\sum_{i=1}^{n} I_{[X_{(ii)}>t]}\right)$ is an unbiased estimator of p with a smaller variance than the empirical survival distribution function based on a simple random sampling of size n. Another unbiased estimator of p in this case is $I_{[X_{(11)}>\frac{t}{n}]}$. So let us again consider the class of estimators of p which are of the form $\left(\sum_{i=1}^{n} c_i I_{[X_i>\alpha t]}\right)$, where $c_i \geq 0, 1 \leq i \leq n \sum_{i=1}^{n} c_i = 1$ and $\alpha > 0$. Clearly, $I_{[X_{(11)}>\frac{t}{n}]}$ and $\frac{1}{n}\left(\sum_{i=1}^{n} I_{[X_{(ii)}>t]}\right)$ are two members in this class which are unbiased estimators of p. It is quite natural to ask the following questions: (i) Can we identify all the unbiased estimators of p in this class? (ii) Can we compare these unbiased estimators based on their variances?

In Section 2, we show that in the case of estimators of p which are of the form

$$\left(\sum_{i=1}^{n} c_i I_{[X_{(ii)}>\alpha t]}\right) \tag{1.1}$$

where $c_i \geq 0,\ 1 \leq i \leq n \quad \sum_{i=1}^{n} c_i = 1$ and $\alpha > 0$, there are only n unbiased estimators of p.

$$(\hat{p})_{1,\cdots,r} = \sum_{i=1}^{r} \left[\frac{\binom{n-i}{n-r}}{\binom{n}{r-1}} \right] I_{[X_{(ii)} > \frac{t}{(n-r+1)}]} \qquad 1 \le r \le n \quad (1.2).$$

Note that $(\hat{p})_{1,\cdots,r}$ depends only on the first r observations of the RSS of the size n. In Section 3, we examine the variances of these unbiased estimators. We show that for each n, the variance of $(\hat{p})_1$ is the largest among the variances of $(\hat{p})_{1,\cdots,r}$, $1 \le r \le n$. We also show that for $1 \le n \le 3$, $Var((\hat{p})_{1,\cdots,r}) \downarrow$ in r, $1 \le r \le n$. We conjecture that for all n, $Var[(\hat{p})_{1,\cdots,n}]$, is the smallest among $Var\ [(\hat{p})_{1,\cdots,r}]$, $1 \le r \le n$. We show that $Var\ [(\hat{p})_{1,\cdots,n}]$ decreases in n.

2. Unbiased estimators of p based on RSS

Recall that $p = e^{-t/\theta}$ is reliability of a component, at an arbitrary but fixed point in time $t > 0$, whose lifetime is represented by an exponential random variable X with unknown mean $\theta > 0$. We wish to estimate p based on $X_{(11)}, X_{(22)}, \cdots, X_{(nn)}$, an RSS of size n, where $X_{(ii)}$ is the i^{th} order statistic in the random sample $X_{(i1)}, X_{(i2)}, \cdots, X_{(in)}$, $1 \le i \le n$, and the n random samples are independent. The observations $X_{(ii)}$, $1 \le i \le n$, of an RSS of size n may be viewed as the lifetimes of independent $(n-i+1)$-out-of-n systems, each consisting of n components whose lifetimes are independent exponentially distributed with mean θ.

The following Theorem, which is the main result of this section, shows that there are exactly n unbiased estimators and given then explicit form.

Theorem 2.1 Let $p = e^{-t/\theta}$ be the reliability of the component whose lifetime X is an exponential random variable with mean $\theta > 0$. Let $X_{(11)}, X_{(22)}, \cdots, X_{(nn)}$ be an RSS of size n. Among all estimators of p which are of the form as in (1.1), the following are the only unbiased estimators of p

$$(\hat{p})_{1,\cdots,r} = \sum_{i=1}^{r} \left[\frac{\binom{n-i}{n-r}}{\binom{n}{r-1}} \right] I_{\left[X_{(ii)} > \frac{t}{(n-r+1)}\right]} \tag{2.1}$$

where $\quad 1 \le r \le n$.

with variance

$$Var\,[(\hat{p})_{1,\cdots,r}] = \sum_{i=1}^{r} \left[\frac{\binom{n-i}{n-r}}{\binom{n}{r-1}} \right]^2 \Pi_i\left(\frac{t}{n-r+1}\right)\left[1 - \Pi_i\left(\frac{t}{n-r+1}\right)\right] \tag{2.2}$$

where for all $1 \le i \le r$ and $\alpha > 0$,

$$\Pi_i(\alpha t) = P[X_{(ii)} > \alpha t] = i \binom{n}{i} \sum_{k=0}^{i-1} (-1)^k \binom{i-1}{k} \int_0^{p^\alpha} v^{n-1+k} dv \tag{2.3}$$

Proof Consider an estimator $\hat{p} = \sum_{i=1}^{n} c_i(\alpha) I_{[X_{(ii)} > \alpha t]}$, where $c_i(\alpha) \ge 0$, $1 \le i \le n$ $\sum_{i=1}^{n} c_i(\alpha) = 1$, and $\alpha > 0$. For this estimator to be an unbiased estimator of $p = e^{-t/\theta}$, we must have

$$\sum_{i=1}^{n} c_i(\alpha)\, P[X_{(ii)} > \alpha t] = p. \tag{2.4}$$

The LHS of (2.4) is expressed as a polynomial in p^{α}, then (2.4) is used to show that the only choices for α, $c_i(\alpha)$ are the ones given in (2.1).

First we use the following well known expression

$$P\,[X_{(ii)} > \alpha t] = i\binom{n}{i} \int_{o}^{p^{\alpha}} v^{n-1}(1-v)^{i-1}\,dv$$

$$= \sum_{k=0}^{i-1} (-1)^k i \binom{n}{i}\binom{i-1}{k} \int_0^{p^{\alpha}} v^{n-i+k}\,dv \qquad (2.5)$$

where $1 \le r \le n$, and $\alpha > 0$. In view of (2.5), the LHS of (2.4) can be written as $\sum_{i=1}^{n} \sum_{k=0}^{i-1} c_i(\alpha)\, i \binom{n}{i} (-1)^k \binom{i-1}{k} \int_0^{p^{\alpha}} v^{n-i+k}\,dv$.

Let $i-k=u$ and making some algebraic manipulations lead to the following form for the LHS of (2.4)

$$\sum_{u=1}^{n}\left[\sum_{k=0}^{n-u} (-1)^k c_{k+u}(\alpha)\binom{n-u}{k}\right]\binom{n}{u-1}(p^{\alpha})^{n-u+1}. \qquad (2.6)$$

In view of (2.6) and let $p^{\alpha} = x$, we can now rewrite (2.4) as

$$\sum_{u=1}^{n}\left[\sum_{k=0}^{n-u} (-1)^k c_{k+u}(\alpha)\binom{n-u}{k}\right]\binom{n}{u-1}(x)^{n-u+1} = x^{1/\alpha} \qquad (2.7)$$

for all $0 < x < 1$. It then follows that $(\frac{1}{\alpha})\,(\frac{1}{\alpha}-1)\cdots(\frac{1}{\alpha}-(n-1))\; x^{(\frac{1}{\alpha}-n)}$ is free of x. Consequently the only possible choices for α are $1, \frac{1}{2}, \cdots, \frac{1}{n}$.

By letting $\alpha = \frac{1}{(n-r+1)}$, $1 \le r \le n$, and equating the coefficient of various powers of x on both sides of (2.7) we get: $c_{r+1}(\alpha)=c_{r+2}(\alpha)=\cdots=c_n(\alpha)=0,\ c_r(\alpha)=\dfrac{1}{\binom{n}{r-1}}$, and

$$c_{r-l}(\alpha)=\sum_{m=1}^{l} c_{r-l+m}(\alpha)\binom{n-r+l}{m}(-1)^{m+1}\,,\ 1\le l\le r-1. \quad (2.8)$$

Finally we show by induction that

$$c_{r-l}(\alpha)=\frac{\binom{n-r+l}{l}}{\binom{n}{r-1}},\quad 0\le l\le r-1. \quad (2.9)$$

Clearly (2.9) is true when $l=0$, Assume (2.9) is true for $c_r(\alpha),\cdots,c_{r-l+1}(\alpha)$, and substituting in (2.8) we get

$$c_{r-l}(\alpha)=\frac{\left[\binom{n-r+l}{l}\left\{\binom{i}{1}-\binom{i}{2}+\cdots+(-1)^{i-1}\binom{i}{1}\right\}\right]}{\binom{n}{r-1}}=\frac{\binom{n-r+l}{l}}{\binom{n}{r-1}}$$

This establishes (2.9) and the Theorem.

We close this section with the following Lemma in the particular case of $n=2$.

Lemma 2.2 Consider an RSS of size $n=2$. Among all convex linear combinations of the form $c_1 I_{[X_{(11)}>\alpha_1 t]}+c_2 I_{[X_{(22)}>\alpha_2 t]}$, the following are the only unbiased estimators of $p=e^{-t/\theta}$

$$(\hat{p})_1=I_{[X_{(11)}>t/2]},\ (\hat{p})_{1,2}=\frac{1}{2}I_{[X_{(11)}>t]}+\frac{1}{2}I_{[X_{(22)}>t]}. \quad (2.10)$$

Proof By equating the expectation of the proposed estimator to p we get

$$c_1 p^{2\alpha_1} + c_2 [p^{2\alpha_2} + p^{2\alpha_2}(1 - p^{\alpha_2})] = p \text{ for all } p. \tag{2.11}$$

equating the second derivative of the LHS of (2.11) to zero we get

$$\begin{aligned} 2c_1 \alpha_1 (2\alpha_1 - 1) p^{2\alpha_1 - 2} + 2c_2 \alpha_2 (\alpha_2 - 1) p^{\alpha_2 - 2} \\ = 2c_2 \alpha_2 (2\alpha_2 - 1) p^{2\alpha_2 - 2}. \end{aligned} \tag{2.12}$$

Substituting x for p^{α_2} in (2.12) we get

$$\begin{aligned} 2c_1 \alpha_1 (2\alpha_1 - 1) x^{2\alpha_1/\alpha_2} + 2c_2 \alpha_2 (\alpha_2 - 1) x \\ = 2c_2 \alpha_2 (2\alpha_2 - 1) x^2 \end{aligned} \quad \text{for all } x. \tag{2.13}$$

Hence we deduce that either $\alpha_1 = \alpha_2 = 1$ and $c_1 = c_2 = \frac{1}{2}$ or $\alpha_1 = \frac{1}{2}$ and $c_1 = 1, c_2 = 0$. This establishes the Lemma.

Note that the above Lemma strengthens Theorem 2.1 in the particular case of $n = 2$.

3. Comparing the variance of the unbiased estimators of p

In this section, the following Theorem shows that $(\hat{p})_1$ has the largest variance among the n unbiased estimators $(\hat{p})_{1,\cdots,r}$, $1 \le r \le n$.

Theorem 3.1 Let $(\hat{p})_{1,\cdots,r}$ be the unbiased estimators of $p = e^{-t/\theta}$ given as

$$(\hat{p})_{1,\cdots,r} = \sum_{i=1}^{r} \left[\frac{\binom{n-i}{n-r}}{\binom{n}{r-1}} \right] I_{[X_{(ii)} > \frac{t}{(n-r+1)}]}, 1 \le r \le n. \tag{3.1}$$

Then

$$Var\,[(\hat{p})_{1,\cdots,r}] \le \frac{(n-r+1)}{n} Var\,[(\hat{p})_1],\ 1 \le r \le n,\ n \ge 1. \tag{3.2}$$

Proof For fixed n, r, $1 \le r \le n$, let us denote $\dfrac{\binom{n-i}{n-r}}{\binom{n}{r-1}}$ by c_i, and

$p\,[X_{(ii)} > \dfrac{t}{(n-r+1)}]$ by p_i, $1 \le r \le n$. Since $(\hat{p})_{1,\cdots,r}$ is an unbiased estimator of p, $1 \le r \le n$, we must have $p = \sum_{i=1}^{r} c_i p_i$, where $0 < c_i < 1, \sum_{i=1}^{r} c_i = 1$, $0 < p_i < 1$, $i = 1,\cdots,r$. This together with the fact that $x(1-x)$ is a concave function of x, $0 < x < 1$, leads to the following inequality

$$p(1-p) \ge \sum_{i=1}^{r} c_i p_i (1-p_i) \tag{3.3}$$

Also note that

$$\sum_{i=1}^{r} c_i p_i (1-p_i) \ge (\max_{1 \le i \le r} c_i) \sum_{i=1}^{r} c_i^2 p_i (1-p_i) = \frac{1}{(\max_{1 \le i \le r} c_i)\, Var\,[(\hat{p})_{1,\cdots,r}]},$$

In view of (3.3) and the fact that $p(1-p) = Var[(\hat{p})_{1,\cdots,r}]$, we get

$$Var\,[(\hat{p})_{1,\cdots,r}] \le (\max_{1 \le i \le r} c_i)\, Var\,[(\hat{p})_1], 1 \le r \le n \tag{3.4}$$

By observing that $\left(\max_{1 \le i \le r} c_i\right) = c_1 = \dfrac{(n-r+1)}{n}$, (3.2) is now established.

Remark 3.2 Let r_n, $n \ge 1$, be a sequence of positive integer such that $n-r+1 = o(n)$, then in view of Theorem 3.1, $(\hat{p})_{1,\cdots,r_n}$ is weakly consistent estimator of p.

In Lemma 3.5 below, we show that for $n=3$, $Var\,[(\hat{p})_1] \geq Var\,[(\hat{p})_{1,2}] \geq Var\,[(\hat{p})_{1,2,3}]$. The proof of this Lemma utilizes the powerful tools of majorization and Schur function, which are widely used in reliability theory, (Marshall and Olkin, 1979). For the sake of completeness we now present the following definitions.

Definition 3.3 Let x and y be two vectors in $R^k, k \geq 2$. Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ and $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$ be increasing arrangements of the vectors x and y respectively. The vector x is said to majorize the vector y, written $x \geq^m y$, if $\sum_{i=1}^{j} x_{(i)} \leq \sum_{i=1}^{j} y_{(i)}, 1 \leq j \leq k-1$ and $\sum_{i=1}^{k} x_{(i)} \leq \sum_{i=1}^{k} y_{(i)}$.

Definition 3.4 A function $f : R^k \to R, k \geq 2$, is said to be Schur convex (Schur concave) if $f(x) \geq f(y)$ $(f(x) \leq f(y))$ whenever $x \geq^m y$.

It is well known that $f(x) = \sum_{i=1}^{k} g(x_i)$ is Schur convex (Schur concave) in x *iff* $g : R \to R$ is convex (concave).

Lemma 3.5 For $n=3$, let $(\hat{p})_1, (\hat{p})_{1,2}$, and $(\hat{p})_{1,2,3}$ be the unbiased estimators of $p = e^{-t/\theta}$, given in (3.1). Then $Var\,[(\hat{p})_1] \geq Var\,[(\hat{p})_{1,2}] \geq Var\,[(\hat{p})_{1,2,3}]$.

Proof We show that $Var\,[(\hat{p})_{1,2}] \geq Var\,[(\hat{p})_{1,2,3}]$,

where

$$(\hat{p})_{1,2} = \frac{2}{3} I_{[X_{(11)} > \frac{t}{2}]} + \frac{1}{3} I_{[X_{(22)} > \frac{t}{2}]}$$

$$(\hat{p})_{1,2,3} = \frac{1}{3}(I_{[X_{(11)}>t]} + I_{[X_{(22)}>t]} + I_{[X_{(33)}>t]}).$$

Let $p_i = P[X_{(ii)} > \frac{t}{2}]$, $i = 1,2$ and $p_i^* = P[X_{(ii)} > t]$, $i = 1,2,3$.

Clearly $p_1 \le p_2$ and $p_1^* \le p_2^* \le p_3^*$. Also since both $(\hat{p})_{1,2}$, and $(\hat{p})_{1,2,3}$, are unbiased estimators of p, we have

$$2p_1 + p_2 = p_1^* + p_2^* + p_3^*. \tag{3.5}$$

Let us now show that $(p_1^*, p_2^*, p_3^*) \ge^m (p_1, p_2, p_3)$. In view of (2.6), we only need to show that $p_1^* \le p_1$ and $p_3^* \ge p_2$.

Clearly $p_1 = P\,[X_{(11)} > \frac{t}{2}] \ge P\,[X_{(11)} > t] = p^*$. Now to see that $p_3^* \ge p_2$, recall that p_3^* may be viewed as the reliability of a 1-out-of-3 system whose independent components have the common reliability p, while p_2 may be viewed as the reliability of a 2-out-of-3 system whose independent components have the common reliability $\sqrt{p}$. The inequality $p_2 \le p_3^*$ follows from the well-known bounds for system reliability given in Theorem 3.4 of Barlow and Proschan (1981). Since $x(1-x)$ is concave function on [0,1], we have

$$\sum_{i=1}^{3} p_i^*(1-p_i^*) \le p_2(1-p_2) + 2p_1(1-p_1) \le p_2(1-p_2) + 4p_1(1-p_1).$$

The Lemma is now established by observing that

$$\sum_{i=1}^{3} p_i^*(1-p_i^*) = 9\ Var\,[(\hat{p})_{1,2,3}],$$

and $p_2(1-p_2) + 4p_1(1-p_1) = 9Var\,[(\hat{p})_{1,2}]$.

It is reasonable to extrapolate from Lemma 3.5 that $\frac{1}{n}\left(\sum_{i=1}^{n} I_{[X_{(ii)}^{n}>t]}\right)$, which is weakly consistent estimator of p, has the minimum variance among $(\hat{p})_{1,2,\cdots r}$, $1 \le r \le n$, $n=1,2,\cdots$. Note that we now add the superscript n in $X_{(ii)}^{n}$ to highlight the dependence of the i^{th} order statistics on the size of the sample n, $1 \le i \le n$.

In the following theorem we show that $Var[(\hat{p})_{1,\cdots n}]$ decreases with n. The proof of this theorem utilizes, again, tools from reliability theory as well as majorization and Schur functions.

Theorem 3.7 Let $V_n = Var\left[\frac{1}{n}\left(\sum I_{[X_{(ii)}^{n}>t]}\right)\right]$, $n \ge 1$. Then

$$\left(\frac{n}{n+1}\right)^2 V_n \le V_{n+1} \le \left(\frac{n}{n+1}\right) V_n. \tag{3.6}$$

Proof Let $p_i = P[X_{(ii)}^{n} > t]$, $1 \le i \le n$ and $p_i^* = P[X_{(ii)}^{n+1} > t]$, $1 \le i \le n+1$. Clearly $p_1 \le p_2 \le \cdots \le n$ and $p_1^* \le p_2^* \le \cdots p_{n+1}^*$. By a well known decomposition of the reliability function of a coherent system, see Lemma 1.1 of Barlow and Proschan (1981), we have

$$p_i^* = p\, p_i + (1-p)\, p_{i-1}, \quad 1 \le i \le n+1, \quad p_0 \equiv 0, \quad p_{n+1} \equiv 1 \tag{3.7}$$

In view of (3.7) and the fact that $p_0 \le p_1 \le \cdots p_{n+1}$, we get

$$p_{i-1} \le p_i^* \le p_i, \qquad 1 \le i \le n+1. \tag{3.8}$$

Next let us show that

$$(n+1)\,(p_r - \overset{*}{p}_r) - r(\overset{*}{p}_{r+1} - \overset{*}{p}_r) = 0, \;\; 1 \le r \le n \tag{3.9}$$

The LHS of (3.9)

$$= (n+1)\left[r\binom{n}{r} \int_0^p v^{n-r}(1-v)^{r-1}\,dv - r\binom{n+1}{r} \int_0^p v^{(n+1)-r}(v-1)^{r-1}\,dv \right]$$

$$- r\left[(r+1)\binom{n+1}{r+1} \int_0^p v^{n-r}(1-v)^{r-1}\,dv - r\binom{n+1}{r} \int_0^p v^{(n+1)-r}(v-1)^{r-1}\,dv \right]$$

$$= \int_0^p r\,v^{n-1}(1-v)^{r-1}\left[\left\{(n+1)\binom{n}{r} - (r+1)\binom{n+1}{r+1}\right\}\right.$$

$$\left. + \left\{ r\binom{n+1}{r} + (r+1)\binom{n+1}{r+1} - (n+1)\binom{n+1}{r} \right\} v \right] dv$$

The integrand above is zero and this establishes (3.9). We now use mathematical induction and (3.9) to show that

$$\begin{aligned}&\left(\sum_{i=1}^{r}(n+1)\,p_i\right) - \left(\sum_{i=1}^{r} n\,\overset{*}{p}_i + r\,\overset{*}{p}_{r+1}\right) \\ &= (n+1)(p_r - \overset{*}{p}_r) - r(\overset{*}{p}_{r+1} - \overset{*}{p}_r) = 0 \;\; 1 \le r \le n-1\end{aligned} \tag{3.10}$$

Clearly (3.10) is true for $r = 1$. Assume (3.10) is true for r, then

$$\begin{aligned}&\left(\sum_{i=1}^{r+1}(n+1)\,p_i\right) - \left(\sum_{i=1}^{r+1} n\,\overset{*}{p}_i + (r+1)\,\overset{*}{p}_{r+2}\right) \\ &= (n+1)\,p_{r+1} - ((n-r)\,\overset{*}{p}_{r+1} + (r+1)\,\overset{*}{p}_{r+2}) \\ &= (n+1)(p_{r+1} - \overset{*}{p}_{r+1}) - (r+1)(\overset{*}{p}_{r+2} - \overset{*}{p}_{r+1}) = 0 \;\;\; 1 \le r \le n-1\end{aligned}$$

Hence (3.10) is true.

We are now ready to show that

$$(\underbrace{\overset{*}{p}_1, \cdots, \overset{*}{p}_1}_{n-times}, \cdots, \underbrace{\overset{*}{p}_{n+1}, \cdots, \overset{*}{p}_{n+1}}_{n-times}) \ge^{m} (\underbrace{p_1, \cdots, p_1}_{(n+1)\,times}, \cdots, \underbrace{p_n, \cdots, p_n}_{(n+1)\,times}). \tag{3.11}$$

We first observe that the coordinates of the two $n(n+1)$ dimensional vectors in (3.11) have the same sum since both $\frac{1}{n}\left(\sum_{i=1}^{n} I_{[X_{(ii)}^{n}>t]}\right)$ and $\frac{1}{n+1}\left(\sum_{i=1}^{n+1} I_{[X_{(ii)}^{n+1}>t]}\right)$ are unbiased estimator of p. Next we need to show that each of the successive partial sums of the vector on the LHS of (3.11) dominates the corresponding partial sum of the vector on the LHS of (3.11). The differences between these partial sums take successively the following forms:

(i) $j\,p_1 - j\,p_1^* \geq 0, 1 \leq j \leq n$ (in view of (3.8));

(ii) $(n+1)\,p_1 - (n\,p_1^* + p_2^* = 0,$ (in view of (3.10));

(iii)

$$\left(\sum_{i=1}^{r}(n+1)\,p_i + l\,p_{r+1}\right) - \left(\sum_{i=1}^{r} n\,p_i^* + (r+l)\,p_{r+1}^*\right) = l\,(p_{r+1} - p_{r+1}^*) \geq 0$$

$1 \leq r \leq n-2,\ 0 \leq l \leq n-r$ (in view of (3.8) and (3.10));

(iv) $$\left(\sum_{i=1}^{r}(n+1)\,p_i + [(n-r)+s]\,p_{r+1}\right) - \left(\sum_{i=1}^{r+1} n\,p_i^* + s\,p_{r+2}^*\right)$$

$$= (n-r)(p_{r+1} - p_{r+1}^*) - s\,(p_{r+2}^* - p_{r+1}^*) \geq 0,\ 1 \leq s \leq r+1$$

(in view of (3.8) and (3.10));

$$\left(\sum_{i=1}^{n-1}(n+1)\,p_i + j\,p_n\right) - \left(\sum_{i=1}^{n} n\,p_i^* + (j-1)\,p_{n+1}^*\right)$$

(v) $= (n(n+1)\,p - (n+1-j)\,p_n) - (n(n+1)\,p - (n+1-j)\,p_{n+1}^*)$

$$= (n+1-j)\,(p_{n+1}^* - p_n) \geq 0 \quad 1 \leq j \leq n+1$$

(in view of (3.8) and the fact that sum of the coordinates of both vectors in(3.11) is equal to $n(n+1)p$). This establishes (3.11).

Now in view of the fact that the function $x(1-x)$ is concave in x, $0 \le x \le 1$. We have $n\left(\sum_{i=1}^{n+1} p_i^* (1-p_i^*)\right) \le (n+1)\left(\sum_{i=1}^{n} p_i\ (1-p_i)\right)$, which leads to the second inequality in (3.6).

We complete the proof by observing that (3.7) together with the concavity of $x(1-x)$ imply that

$$p_i^* (1-p_i^*) \ge p\, p_i (1-p_i) + (1-p)\, p_{i-1} (1-p_{i-1}),\ 1 \le i \le n+1, \qquad (3.12)$$

By summing both side of the inequality in (3.12) we get

$\sum_{i=1}^{n+1} p_i^* (1-p_i^*) \ge \sum_{i=1}^{n} p_i (1-p_i)$. Hence $(n+1)^2 V_{n+1} \ge n^2 V_n$.

In the following Ghitany (2005) shows that $Var\,[(\hat{p})_1] \ge Var\,[(\hat{p})_{1,2}] \ge Var\,[(\hat{p})_{1,2,3}]$, such conjecture is incorrect by considering the case $n = 4,5$ respectively.

Lemma 3.6 For $n=4$, let $(\hat{p})_{1,2,3}$ and $(\hat{p})_{1,2,3,4}$ be two unbiased estimators of $p = e^{-t/\theta}$, $t, \theta > 0$, given by (2.1). Then

$Var\,[(\hat{p})_{1,2,3}] < (=)(>)\ Var\,[(\hat{p})_{1,2,3,4}]$ if $p < (=)(>) p_0$ where $p_0 = 0.078484$.

Proof For $n=4$, using (2.1), the two unbiased estimators $(\hat{p})_{1,2,3}$ and $(\hat{p})_{1,2,3,4}$ of p are given by

$$(\hat{p})_{1,2,3} = \frac{1}{2} I_{[X_{(11)} > (1/2)t]} + \frac{1}{3} I_{[X_{(22)} > (1/2)t]} + \frac{1}{6} I_{[X_{(33)} > (1/2)t]}$$

$$(\hat{p})_{1,2,3,4} = \frac{1}{4}\{I_{[X_{(11)}>t]} + I_{[X_{(22)}>t]} + I_{[X_{(33)}>t]}\}$$

Using (2.2) and (2.3), the variance of these estimator are given by

$$Var\,[(\hat{p})_{1,2,3}] = \frac{1}{6}p + \frac{2}{9}p^{3/2} - p^2 + \frac{8}{3}p^{5/2} - \frac{41}{9}p^3 + 4p^{7/2} - \frac{3}{2}p^4$$

$$Var\,[(\hat{p})_{1,2,3,4}] = \frac{1}{4}p - p^2 + 3p^3 - \frac{13}{2}p^4 + \frac{19}{2}p^5 - 9p^6 + 5p^7 - \frac{5}{4}p^8$$

Let $\xi(p) = Var[(\hat{p})_{1,2,3}] - Var[(\hat{p})_{1,2,3,4}]$. That is,

$$\xi(p) = \frac{1}{36}p(-3 + 8p^{1/2} + 96p^{3/2} - 272p^2 + 144p^{5/2} + 180p^3 - 342p^4$$
$$+324p^5 - 180p^6 + 45p^7)$$

The first and second derivatives of $\xi(p)$, respectively, are given by

$$\xi'(p) = -\frac{1}{12} + \frac{1}{3}p^{1/2} + \frac{20}{3}p^{3/2} - \frac{68}{3}p^2 + 14p^{5/2} + 20p^3 - \frac{95}{2}p^4$$
$$+54p^5 - 35p^6 + 10p^7),$$

$$\xi''(p) = \frac{1}{6}p^{-1/2} + 10p^{1/2} - \frac{136}{3}p + 35p^{3/2} + 60p^2 - 190p^3 + 270p^4$$
$$-210p^5 + 70p^6$$

The function $\xi(p)$ has two critical points at $p_1 = 0.035744$ and $p_2 = 0.651382$. Since $\xi'(p_1) = 1.456697$ and $\xi''(p_2) = -0.577510$, then p_1 is the point of minima and p_2 is the point of maxima. Also, since $\xi(0) = \xi(1) = 0$, the function $\xi(p)$ decreases on $[0, p_1]$, increases on $[p_1, p_2]$, and decreases on $[p_2, 1]$. Hence there exist a unique point $p_0 = 0.078484 \in (p_1, p_2)$ such that $\xi(p) < (=)(>) 0$ for $p < (=) > p_0$.

Finally, we consider the case when the sample size is five. The proof of the following Lemma is similar to Lemma 3.6 above and hence is omitted.

Lemma 3.7 For $n=5$, let $(\hat{p})_{1,2,3,4}$ and $(\hat{p})_{1,2,3,4,5}$ be two unbiased estimators of $p=e^{-t/\theta}$, $t,\theta>0$, given by (2.1). Then

$$Var\,[(\hat{p})_{1,2,3,4}] < (=)\,(>)\; Var\,[(\hat{p})_{1,2,3,4,5}] \text{ if } p < (=)\,(>)\; p_0^*$$

where $p_0^* = 0.116044$.

CHAPTER IV

HYPOTHESIS TESTING BASED ON RANKED SET SAMPLING

1. Introduction

In this chapter, we assume that $F(x)$ is normal with unknown mean μ and known variance (=1, without loss of generality), and address the problem of testing the hypothesis $H_0 : \mu = 0$ verses $H_1 : \mu > 0$. Based on RSS and its various modifications, as introduced in Sinha *et al.* (1996), we propose a variety of exact tests for the above problem and point out that there do exist many test with much better power properties compared to the normal test based on SRS.

In Section 2, test on H_0 based on $\hat{\mu}_{rss}$, $\hat{\mu}_{blue}$, $\hat{\mu}_{prss}(l)$, $\hat{\mu}_{median:n}(l)$, and $\hat{\mu}_{smallest}(l)$, are discussed. Recall that the usual normal test for H_0 based on SRS of size n reject H_0 if

$$\sqrt{n}\,\overline{X} > z_\alpha \tag{1.1}$$

with its power at $\mu > 0$ as

$$Power\ (\mu | SRS) = 1 - \Phi\ (z_\alpha - \sqrt{n}\ \mu) \tag{1.2}$$

where z_α is the power α point of standard normal distribution and $\Phi(.)$ is the standard normal *cdf*.

In Section 3, we obtained the UMPT for one-sided alternative and the LRT for the two-sided alternative in case of the exponential distribution and the UMPT for the two-sided alternative in case of uniform distribution using SRS and will adapt these tests to RSS.

2. Improved test of mean of normal $N(\mu, 1)$ distribution based on RSS and its modification

2.1 Test based on McIntyre's $\hat{\mu}_{rss}$

Since the statistic $\hat{\mu}_{rss}$, is unbiased, we propose a test statistic based on $\hat{\mu}_{rss}$, for H_0 and reject H_0 if

$$\hat{\mu}_{rss}, > c_{1,\alpha}, \tag{2.1}$$

with its power at $\mu > 0$ as

$$Power\ (\mu | SRS) = P\{\hat{\mu}_{rss} > c_{1,\alpha} \mid \mu\}, \tag{2.2}$$

where $c_{1,\alpha}$ is the upper α level cut-off point of $\hat{\mu}_{rss}$, under H_0.

2.2 Test based on RSS BLUE $\hat{\mu}_{blue}$

Using the fact that $E\ (X_{(ii)}) = \mu + v_i\,\sigma$ and var $(X_{(ii)}) = \gamma_i\,\sigma$ where v_i and γ_i are respectively the expected value and the variance of the i^{th} order statistic in a sample of size n from a standard normal population, Sinha *et al.* (1996), derive the BLUE of μ based on RSS as

$$\hat{\mu}_{blue} = \frac{\sum_{i=1}^{n} \frac{X_{(ii)}}{\gamma_i}}{\sum_{i=1}^{n} \frac{1}{\gamma_i}} \tag{2.3}$$

with

$$\text{var}(\hat{\mu}_{blue}) = \frac{\sigma^2}{\sum_{i=1}^{n} \frac{1}{\gamma_i}} \tag{2.4}$$

and as shown that $\operatorname{var}(\hat{\mu}_{blue})$, which is always smaller than $\operatorname{var}(\hat{\mu}_{rss})$, is indeed much smaller than $\operatorname{var}(\overline{X})$, (see chapter I).

We propose a test for H_0 based on $\hat{\mu}_{blue}$, Reject H_0 if

$$\hat{\mu}_{blue} > c_{2,\alpha} \tag{2.5}$$

with its power at $\mu > 0$ as

$$Power\ (\mu \mid \hat{\mu}_{blue}) = P\{\hat{\mu}_{blue} > c_{2,\alpha} \mid \mu\} \tag{2.6}$$

where $c_{2,\alpha}$ is the upper α level cut-off point of $\hat{\mu}_{blue}$, under H_0.

2.3 Test based on PRSS BLUE $\hat{\mu}_{prss}(l)$

Sinha *et al.* (1996) also derived the BLUE of μ based partial RSS, as

$$\tilde{\mu}_{prss}(l) = \frac{\left(\sum_{i=1}^{l}\frac{v_i^2}{\gamma_i}\right)\left(\sum_{i=1}^{l}\frac{X_{(ii)}}{\gamma_i}\right) - \left(\sum_{i=1}^{l}\frac{v_i}{\gamma_i}\right)\left(\sum_{i=1}^{l}X_{(ii)}\frac{v_i}{\gamma_i}\right)}{\left(\sum_{i=1}^{l}\frac{v_i^2}{\gamma_i}\right)\left(\sum_{i=1}^{l}\frac{1}{\gamma_i}\right) - \left(\sum_{i=1}^{l}\frac{v_i}{\gamma_i}\right)^2} \tag{2.7}$$

with

$$\operatorname{var}(\tilde{\mu}_{prss}(l)) = \frac{\sigma^2\left(\sum_{i=1}^{l}\frac{v_i^2}{\gamma_i}\right)}{\left(\sum_{i=1}^{l}\frac{v_i^2}{\gamma_i}\right)\left(\sum_{i=1}^{l}\frac{1}{\gamma_i}\right) - \left(\sum_{i=1}^{l}\frac{v_i}{\gamma_i}\right)^2} \tag{2.8}$$

and showed that often a partial RSS combined with optimum weight as above does better than a SRS of size n. $\hat{\mu}_{prss}(l)$ has the added advantage of using a much smaller number of actual measurements.

We therefore propose a test procedure for H_0 based on $\hat{\mu}_{prss}$, Reject H_0 if

$$\hat{\mu}_{prss}(l) > c_{3,\alpha}(l), \qquad l = 2,3,\cdots,n-1 \tag{2.9}$$

with its power at $\mu > 0$ as

$$Power\ (\mu \mid \hat{\mu}_{prss}(l)) = P\{\hat{\mu}_{prss}(l) > c_{3,\alpha}(l) \mid \mu\}, \quad l = 2,3,\cdots,n-1 \tag{2.10}$$

where $c_{3,\alpha}$ is the upper α level cut-off point of $\hat{\mu}_{prss},(l)$ under H_0 .

2.4 Test based on sample medians

To address the question of right selection of order statistics from Table 1.1, Sinha *et al.* (1996) noted the following variance inequality for order statistics of a normal distribution.

Lemma 2.1 var $(X_{median:n}) \leq$ var $(X_{r:n})$ for any r and n.

In view of the above result, Sinha *et al.* (1996) recommended the use of the median from each set of n observations, and the mean of all such medians as an estimator of μ, namely,

$$\hat{\mu}\ (n:n) = \frac{1}{n}[X^{(1)}_{median:n} + \cdots + X^{(n)}_{median:n}]$$

where $X^{(i)}_{median:n}$ is the sample median from the i^{th} row of the Table 1.1.

Slightly more efficiently, Sinha *et al.* (1996) proposed measuring only m medians from the first m rows of Table 1.1, where $m \leq n$, and use

$$\hat{\mu}\ (m:n) = \frac{1}{m}[X^{(1)}_{median:n} + \cdots + X^{(m)}_{median:n}] \tag{2.11}$$

as an estimator of μ. They also proved that $m = 2$ is enough to dominate $\overline{X}$ because of the following powerful and exact result.

Theorem 2.1 $\text{var}(\bar{X}_n) < \text{var}(X_{median:n}) < 2\,\text{var}(\bar{X}_n)$

Based on above considerations for n *odd* we proposed to use $\hat{\mu}_{median:n}(l) = \hat{\mu}\,(l,n)$, the average of $l \le n$ medians, as our test statistic, Reject H_o if

$$\hat{\mu}_{median:n}(l) > c_{4,\alpha}(l), \qquad l = 2,3,\cdots,n \qquad (2.12)$$

with its power at $\mu > 0$ as

$$Power\ (\mu|\,\hat{\mu}_{median:n}(l)) = P\{\hat{\mu}_{median:n}(l) > c_{4,\alpha}(l)|\,\mu\}, l = 2,3,\cdots,n \quad (2.13)$$

where $c_{4,\alpha}$ is the upper α level cut-off point of $\hat{\mu}_{median:n}(l)$ under H_o.

If n is even ($=2k$, say), combine in pairs the k^{th} order statistic from one row with the $(k+1)st$ order statistic of the next row, and define the unbiased estimator of μ as

$$\hat{\mu}^{*}_{median:n}(l) = \frac{1}{l}[(X_{(1k)} + X_{(2(k+1))}) + (X_{(3k)} + X_{(4(k+1))}) + \cdots$$
$$+ (X_{((l-1)k)} + X_{(l(k+1))})],\ l = 2,4,6\cdots n \qquad (2.14)$$

then propose a test procedure for H_o based on $\hat{\mu}^{*}_{median:n}(l)$, Reject H_o if

$$\hat{\mu}^{*}_{median:n}(l) > c^{*}_{4,\alpha}(l), \qquad l = 2,4,6,\cdots,n \qquad (2.15)$$

with its power at $\mu > 0$ as

$$Power\ (\mu\Big|\,\hat{\mu}^{*}_{median:n}(l)) = P\{\hat{\mu}^{*}_{median:n}(l) > c^{*}_{4,\alpha}(l)|\,\mu\}, \qquad (2.16)$$
$$l = 2,4,6,\cdots,n$$

where $c_{4,\alpha}^{*}$ is the upper α level cut-off point of $\hat{\mu}_{median:n}^{*}(l)$ under H_o.

2.5 Test based on $\tilde{\mu}_{smallest}(l)$

Sinha *et al.* proposed $\tilde{\mu}_{\min} = \frac{1}{n}\sum_{i=1}^{n} X_{(i1)} - \nu_1$, bias corrected mean of n smallest observations as an estimator of μ, and, more efficiently, suggested the use of

$$\tilde{\mu}_{\min}(m) = \frac{1}{m}\sum_{i=1}^{m} X_{(i1)} - \nu_1 \qquad (2.17)$$

based on $m(< n)$ smallest means.

We define

$$\tilde{\mu}_{smallest}(l) = \frac{1}{l}\sum_{i=1}^{l} X_{(i1)}, \qquad l = 2,3,\cdots,n \qquad (2.18)$$

and propose a test procedure for H_o based on $\tilde{\mu}_{smallest}(l)$, Reject H_o if

$$\tilde{\mu}_{smallest}(l) > c_{5,\alpha}(l), \qquad l = 2,3,\cdots,n \qquad (2.19)$$

with its power at $\mu > 0$ as

$$Power\ (\mu | \tilde{\mu}_{smallest}(l)) = P\{\tilde{\mu}_{smallest}(l) > c_{5,\alpha}(l) | \mu\}, l = 2,3,\cdots,n \quad (2.20)$$

where $c_{5,\alpha}$ is the upper α level cut-off point of $\tilde{\mu}_{smallest}(l)$ under H_o.

3. Tests for exponential distribution

Let $X_1, X_2, \cdots, X_n$ be a random sample from the exponential distribution with *pdf*

$$f(x)=\begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}} & \text{if } x>0 \\ 0 & \text{otherwise} \end{cases}$$

We are interested in testing the hypothesis

$$H_0:\theta=\theta_0 \quad vs \qquad H_\alpha:\theta>\theta_0 \tag{3.1}$$

It is well known that the UMPT of size α for testing (3.1) is given by

$$\phi UMPT=\begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i > C_\alpha \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

without loss of generality we may take $\theta_0=1$. Then $C_\alpha=\dfrac{\chi^2_{2n,1-\alpha}}{2}$, where χ^2_m is the *chi-square* distribution with m degree of freedom. The power of test (3.2) is given by

$$\beta_{\phi UMPT}(\theta)=P_\theta\left(\sum_{i=1}^{n} X_i>\frac{1}{2}\chi^2_{2n,1-\alpha}\right)=P_\theta\left(W>\frac{1}{\theta}\chi^2_{2n,1-\alpha}\right),$$

where W is distributed χ^2_{2n}.

To obtain the test using RSS, namely, $X_{(11)},X_{(22)},\cdots,X_{(nn)}$ be the n independent random variables all with the same *cdf* $F(x)$. To simplify the notation, $X_{(ii)}$ will be denoted by Y_i through out this section.

To test the same hypothesis (3.1) using RSS, the test statistic is given by

$$\phi_1=\begin{cases} 1 & \text{if } \sum_{i=1}^{n} Y_i > d \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where d is determined so that the test ϕ_1 has size α. To obtain the value of d, we need the distribution of $\sum_{i=1}^{n} Y_i$ under H_0. For this, consider the following transformation

$$Z_1 = Y_1,\ Z_2 = Y_1 + Y_2,\ Z_3 = Y_1 + Y_2 + Y_3, \cdots, Z_n = \sum_{i=1}^{n} Y_i.$$

The joint *pdf* of $Y_1, Y_2, \cdots, Y_n$ are given by

$$g_\theta\,(y_1, \cdots y_n) = \begin{cases} \prod_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!}[1-e^{-\frac{y_i}{\theta}}]^{i-1}\}\frac{1}{\theta^n}\exp(-\frac{1}{\theta}\sum_{i=1}^{n}(n-i+1)y_i) \\ \qquad\qquad y_i > 0, \qquad i = 1, \cdots, n \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

(3.4)

Then the joint *pdf* of $Z_1, Z_2, \cdots, Z_n$ is given by

$$h_\theta\,(z_1, z_2, \cdots, z_n) = g_\theta\,(z_1, z_2 - z_1, z_3 - z_2, \cdots, z_n - z_{n-1})$$

which implies that the *pdf* of Z_n is

$$k_\theta(z_n) = \int_0^{z_n}\int_0^{z_{n-1}}\cdots\int_0^{z_2} g_\theta(z_1, z_2 - z_1, \cdots z_n - z_{n-1})\ dz_1\, dz_2 \cdots dz_{n-1}$$

(3.5)

Therefore the power function of the test (3.3) is given by

$$\beta_{\phi_1}(\theta) = P_\theta(\sum_{i=1}^{n} Y_i > d) = \int_d^{\infty} k_\theta(z_n)\, dz_n,$$

To find d, we need to solve

$$\beta_{\phi_1}(1) = \alpha = \int_d^{\infty} k_{\theta=1}(z_n)\, dz_n. \qquad (3.6)$$

Next we will consider the LRT for testing the hypothesis

$$H_0:\theta=1 \quad vs \quad H_\alpha:\theta\neq 1 \tag{3.7}$$

It is well known that the LRT of size α is given by

$$\phi_{LRT}=\begin{cases}0 & if\ \dfrac{\chi^2_{2n,\alpha/2}}{2}<\sum_{i=1}^{n}X_i<\dfrac{\chi^2_{2n,1-(\alpha/2)}}{2}\\ 1 & otherwise\end{cases} \tag{3.8}$$

and its power function is given by

$$\begin{aligned}\beta_{\phi\ LRT}(\theta)&=1-P_\theta\left(\frac{\chi^2_{2n,\alpha/2}}{2}<\sum_{i=1}^{n}X_i<\frac{\chi^2_{2n,1-(\alpha/2)}}{2}\right)\\ &=1-P_\theta\left(\frac{\chi^2_{2n,\alpha/2}}{\theta}<W<\frac{\chi^2_{2n,1-(\alpha/2)}}{\theta}\right)\end{aligned}$$

where $W=\sum_{i=1}^{n}X_i$ is distributed as χ^2_{2n}.

To test the same hypothesis using RSS, the following test is proposed

$$\phi_2=\begin{cases}0 & if\ k_1<\sum_{i=1}^{n}Y_i<k_2\\ 1 & otherwise\end{cases}$$

The power function of the test ϕ_2 is

$$\beta_{\phi_2}(\theta)=1-P_\theta\left(k_1<\sum_{i=1}^{n}Y_i<k_2\right)=1-\int_{k_1}^{k_2}k_\theta(z_n)\,dz_n$$

where $k_\theta(z_n)$ is defined in (3.5). To obtain the test of size α we need to find k_1 and k_2 satisfy

$$\beta_{\phi_2}(1)=\alpha=1-\int_{k_1}^{k_2}k_{\theta=1}(z_n)\,dz_n.$$

We will take $1-\int_0^{k_1} k_{\theta=1}(z_n)\,dz_n = \frac{\alpha}{2}$ and $1-\int_0^{k_2} k_{\theta=1}(z_n)\,dz_n = 1-\frac{\alpha}{2}$.

4. Tests for uniform distribution

Let $X_1, X_2, \cdots, X_n$ be a random sample from the uniform distribution with probability density function

$$f(x)=\begin{cases}\dfrac{1}{\theta} & \text{if } 0\le x\le \theta\\ 0 & \text{otherwise}\end{cases}$$

We are interested in testing the hypothesis

$$H_0:\theta=\theta_0 \quad vs \quad H_\alpha:\theta\ne\theta_0 \tag{4.1}$$

without loss of generality we may take $\theta_0=1$. Since the UMPT test for (4.1) exists there is no need to consider the LRT. The UMPT of size α is given by

$$\phi=\begin{cases}1 & \text{if } X_{(n)}>1 \text{ or } X_{(n)}\le n\sqrt{\alpha}\\ 0 & \text{otherwise}\end{cases} \tag{4.2}$$

where $X_{(n)}$ is the largest ordered statistic. Then the power function of this test is

$$\phi(\theta)=\begin{cases}1 & \text{if } \theta\le n\sqrt{\alpha}\\ \dfrac{\alpha}{\theta} & \text{if } n\sqrt{\alpha}<\theta\le 1\\ 1+\dfrac{\alpha-1}{\theta} & \text{if } \theta>1\end{cases}$$

To test the same hypothesis using the RSS the test statistic is

$$\phi_3=\begin{cases}1 & \text{if } \max\{Y_i\}<c \text{ or } \max\{Y_i\}>1\\ 0 & \text{otherwise}\end{cases}$$

To find the value of c we solve the equation

$$\alpha = P_{\theta=1}(\max\{Y_i\} < c) = \prod_{i=1}^{n}(P_{\theta=1}(Y_i) < c)$$

which can be written as

$$\alpha = \prod_{i=1}^{n}\left(\int_0^c \frac{n!}{(i-1)!(n-i)} (\frac{y_i}{\theta})^{i-1} (1 - \frac{y_i}{\theta})^{n-i} \frac{1}{\theta}\, dy_i \right).$$

Then the power of the test can be written as

$$\beta_{\phi_3}(\theta) = \begin{cases} 1 & \text{if } \theta \le c \\ \prod_{i=1}^{n} P_\theta(Y_i \le c) & \text{if } c < \theta \le 1 \\ \prod_{i=1}^{n} P_\theta(Y_i \le c) + 1 - \prod_{i=1}^{n} P_\theta(Y_i \le 1) & \text{if } \theta > 1 \end{cases}$$

CHAPTER V

RANKED SET SAMPLING WITH CONCOMITANT VARIABLE

1. Introduction

In this chapter, we concentrate on consistent ranking mechanisms, which is referred to as multi-layer RSS, as discussed in Section 2. The multi-layer ranking mechanism is conceptually equivalent to a stratification of the space of the concomitant variables. The multi-layer RSS is particularly useful for the estimation of the regression coefficients. In Section 3, the estimation of the regression model and regression estimates of the mean of the response variable in the context of RSS are discussed. It is argued that, for the estimation of the mean of the response variable, the RSS regression estimate is better than the RSS sample mean as long as the response variable and the concomitant variables are moderately correlated.

1.2 Use of a concomitant variable

Extending the work of Dell and Clutter (1972) and David and Levine (1972) on ranking errors, Stokes (1977) proposed the use of a concomitant variable. It is supposed that a covariate Y is available which can be easily measured and which is correlated with the variable of interest X. All units are quantified and ranked with respect to Y, only those units selected according to the RSS protocol are quantified with respect to X.

In order to investigate the effect of the concomitant variable Y on the performance of RSS, Stokes made two very restrictive assumptions: (i) the regression of X on Y is linear, and (ii) the standardized variables

$(X-\mu_X)/\sigma_X$ and $(Y-\mu_Y)/\sigma_Y$ follow the same distribution. Notice that both requirements are met when (X, Y) has a bivariate normal distribution.

Let $X_{(11)},\cdots,X_{(nn)}$ denote the ranked set sample of size n drawn with the help of the concomitant variable Y. Since the linear regression relationship is maintained when we ranked according to the concomitant, the assumption (i) implies that

$$X_{(i:n)} = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(Y_{(i:n)} - \mu_Y) \tag{1.1}$$

where the $Y_{(i:n)}$ are the correctly ranked order statistics on Y. Equation (1.1) implies the fundamental relationship

$$\frac{EX_{(i:n)} - \mu_X}{\sigma_X} = \rho\,\frac{EY_{(i:n)} - \mu_Y}{\sigma_Y} \tag{1.2}$$

The relative precision of ranked set sampling for estimating μ_X when ranking is by the concomitant Y (Stokes (1977) and Ridout and Cobby (1987)) is given by

$$RP(X:Y) = \frac{1}{1-\dfrac{\rho^2}{n\sigma_Y^2}\displaystyle\sum_{i=1}^{n}(EY_{(i:n)} - \mu_Y)^2} \tag{1.3}$$

where ρ is the coefficient correlation between X and Y.

2. Multi-layer ranked set sampling

In this Section, a sampling scheme using multiple concomitant variables that will be referred to as a multi-layer ranked set sampling (MRSS). To distinguish, an RSS using a single concomitant variable as the ranking criterion is to be called as marginal RSS. We can represent the population by the sample space of the multiple concomitant variables. In RSS, to rank the sampling units with respect to a singe concomitant variable is like partitioning the space into slices along one axis of the space. If every concomitant variable is reasonably associated with the response variable, the features of Y within each slice are still subject to the variation due to other concomitant variables and hence will not be as homogenous as desired. An obvious alternative is then to partition the space into cubes along all axis. Then we can expect more homogeneity within a cube than within a slice. This leads us to the MRSS scheme to be described in the following

Without loss of generality, consider two concomitant variables for the sake of convenience. Let $X^{\{1\}}$ and $X^{\{2\}}$ denote the concomitant variables. Let n, l be positive integers. A two-layer RSS procedure goes as follows. First, nl^2 independent sets, each of size n, are drawn from population. The units in each of these sets are ranked according to $X^{\{1\}}$. Then, for l^2 ranked sets, the units with $X^{\{1\}}$ rank-1 are selected, for another l^2 ranked sets, the units with $X^{\{1\}}$ rank-2 are selected, and so on. Let the values of $(Y, X^{\{1\}}, X^{\{2\}})$ these selected units be denoted by

$$\begin{array}{c}(Y_{(11)}, X^{\{1\}}_{(11)}, X^{\{2\}}_{(11)}) \cdots (Y_{(1l^2)}, X^{\{1\}}_{(1l^2)}, X^{\{2\}}_{(1l^2)}) \\ \cdots \qquad \cdots \qquad \cdots \\ (Y_{(n1)}, X^{\{1\}}_{(n1)}, X^{\{2\}}_{(n1)}) \cdots (Y_{(nl^2)}, X^{\{1\}}_{(nl^2)}, X^{\{2\}}_{(nl^2)})\end{array} \tag{2.1}$$

where the values of $X^{\{1\}}_{(ri)}, X^{\{2\}}_{(ri)}$ are measured and $Y_{(ri)}$ are latent. This completes the first layer of the procedure. In the second layer, the units represented in each row of (2.1) are divided -randomly or systematically – into l subsets, each of size l. The units in each of these subsets are ranked according to $X^{\{2\}}$. Then, for the first ranked subset, the units with $X^{\{2\}}$-rank 1 are selected and its values on Y is measured, for the second ranked subset, the unit with $X^{\{2\}}$-rank 2 is selected and its value on Y is measured. and so on. This completes one cycle of the procedure. Repeating the cycle m times then yields the data set

$$\{Y_{(r\,s)\,j} : r = 1,\cdots,k; s = 1,\cdots,l; j = 1,\cdots,m\}$$

where $Y_{(r\,s)\,j}$ is the measurement of Y in the j^{th} cycle on the unit with $X^{\{1\}}$- rank r and $X^{\{2\}}$- rank s.

Note that each observation above is generated from nl simple random units independent of others. The procedure described above can be extended to general p- layer RSS straightforwardly only with increasing complexity of notations.

2.1 Consistency of multi-layer ranked set sampling

Let F denote the distribution function of Y, $F_{(r\,s)}$ denote the common distribution function of $Y_{(r\,s)\,j}$ and $F_{(r)}$ denote the common distribution function of the latent values of $Y_{(r)\,i}$. The observation generating procedure of two-layer RSS can be illustrated as follows

$$
Y \sim F \begin{cases}
Y_{(1)} \sim F_{(1)} \begin{cases} Y_{(11)} \sim F_{(11)} \\ Y_{(12)} \sim F_{(12)} \\ \quad \cdots \\ Y_{(1l)} \sim F_{(1l)} \end{cases} \\
Y_{(2)} \sim F_{(2)} \begin{cases} Y_{(21)} \sim F_{(21)} \\ Y_{(22)} \sim F_{(22)} \\ \quad \cdots \\ Y_{(2l)} \sim F_{(2l)} \end{cases} \\
\cdots \\
Y_{(n)} \sim F_{(n)} \begin{cases} Y_{(n1)} \sim F_{(n1)} \\ Y_{(n2)} \sim F_{(n2)} \\ \quad \cdots \\ Y_{(nl)} \sim F_{(nl)} \end{cases}
\end{cases}
$$

Note that $F_{(r)}$ is the distribution function of the r^{th} $X^{\{1\}}$-induced order statistics of a random sample size n from F, and that, for fixed r, $F_{(r\,s)}$ is the distribution function of the s^{th} $X_{(r)}^{\{2\}}$-induced order statistic of a simple random sample of size l from $F_{(r)}$, where $X_{(r)}^{\{2\}}$ is the r^{th} $X^{\{1\}}$-induced order statistic of a random sample size n from $X^{\{2\}}$. It follows from the consistency of a marginal RSS that

$$
F = \frac{1}{n}\sum_{r=1}^{n} F_{(r)}, \qquad F_{(r)} = \frac{1}{l}\sum_{s=1}^{l} F_{(r\,s)},
$$

Therefore

$$
F = \frac{1}{nl}\sum_{r=1}^{n}\sum_{s=1}^{l} F_{(r\,s)}. \tag{2.2}
$$

That is, two-layer RSS is consistent. In general, we can show by the similar argument that any multi-layer RSS is consistent.

2.2 Issues on the choice of concomitant variables

When there are many concomitant variables, we need to make a selection as to which ones are to be used in the MRSS. We make a brief discussion on this issue.

The principle for the selection of concomitant variable is similar to that for the choice of independent variables in the multiple regressions, only variables that are significantly correlated with the variable of interest should be selected variables, which are selected, should not exhibit strong co-linearity. . But, by the nature of the usage of the concomitant variable in RSS, the implementation of the selection differs from that in multiple regression-a choice of the concomitant variables must be made before measurements on the response variable are taken. In the following, we distinguish the selection based on the significance of the concomitant variables and that based on the collinearity among the concomitant variables.

Preliminaries selection based on co-linearity:

The co-linearity among the concomitant variables can be analyzed using the information the concomitant variables alone. Since, by assumption, the values of the concomitant variables can be easily obtained with negligible cost, a preliminary sampling on the values of the concomitant variables can be conducted and the co-linearity among them can then be analyzed. When several variables exhibit a strong co-linearity, only one of them should be selected for the MRSS procedure. Consider an extreme case. Suppose two concomitant variables are almost perfectly correlated. The question whether to select only one of them or not is then equivalent to whether to rank the sampling units based on a single variable in the following way or an other (i) rank of set of n^2 sampling units all together according to the values of the variable, (ii) rank the n^2 units in the manner of a two-layer RSS but use the same variable as the ranking variable at both layers. Suppose the ranking variable is the response variable itself and we are concerned with the estimation of the mean of this variable. The first way of ranking then results in a relative efficiency, roughly, $(n^2+1)/2$. The second way of ranking, however, results in relative efficiency, roughly, $(n+1)^2/4$. The first way of ranking is more always more efficient than the second way of ranking.

Selection based on the significance of the concomitant variables:

Before any measurements on the response variable are taken, the assessment of the significance of a concomitant variable in its relation to the response variable might be made by using some prior information, if available, or by guess work. An objective assessment of the significance can be made after some measurements on the response variable are obtained. The following dynamic version of the multi-layer RSS can be applied in practice. In the earlier stage of the sampling, more concomitant variables, which are potentially significant, are used in the ranking mechanism. Once enough measurements on the response variable are made, the significance of each concomitant variable is assessed. Then, discard the non-significant concomitant variables and use the remaining ones in the ranking mechanism. As long as the ranking mechanism is the same at each cycle of RSS, the resultant sampling is at least more efficient than SRS.

3. Regression analysis based on RSS with concomitant variables

Suppose that the variable Y and concomitant variable $\mathbf{X}$ follows a linear regression model, *i.e.*,

$$Y = \alpha + \boldsymbol{\beta}'\mathbf{X} + \varepsilon \tag{3.1}$$

where $\boldsymbol{\beta}$ is a vector of unknown constant coefficients, and ε is a random variable with mean zero and variance σ_ε^2 and is independent of $\mathbf{X}$.

Suppose that an RSS with certain ranking mechanism is implemented in m cycles. In a typical cycle i, for $r = 1,\cdots,n$, a simple random sample of n units with latent values $(Y_{(1r)i}, \mathbf{X}_{(1r)i}),\cdots,(Y_{(nr)i}, \mathbf{X}_{(nr)i})$ is drawn from the population. The values of the $\mathbf{X}$'s are all measured. The n sampled units are ranked according to the ranking mechanism. Then the Y value of the unit with rank r is measured. The $\mathbf{X}$ and Y values of the unit with rank r are denoted, respectively, by $\mathbf{X}_{(r)i}$ and $Y_{(r)i}$. In the completion of the sampling, we have a data set as follows

$$\begin{array}{c} Y_{(1)i}, \mathbf{X}_{(1)i}, \mathbf{X}_{(11)i}, \cdots \mathbf{X}_{(1n)i}, \\ \cdots \\ Y_{(ni}, \mathbf{X}_{(n)i}, \mathbf{X}_{(n1)i}, \cdots \mathbf{X}_{(nn)i}, \\ i = 1,\cdots,m \end{array} \tag{3.2}$$

Based on the data set (3.2), we consider in this section two problems; (i) the estimation of the regression coefficients, (ii) the estimation of μ_Y, the mean of Y.

3.1 Estimation of regression coefficients with RSS

Denote by $G(y|x)$ the conditional distribution of Y given $\mathbf{X}$. Let $\{(Y_1, \mathbf{X}_1),\cdots,(Y_n, \mathbf{X}_n)\}$ be a simple random sample from the joint distribution of Y and $\mathbf{X}$. Suppose $(R_1,\cdots,R_n)$ is a random permutation of $(1,\cdots,n)$ determined only by $(\mathbf{X}_1,\cdots,\mathbf{X}_n)$.

Lemma 1. The random permutation induced statistics $Y_{R_1},\cdots,Y_{R_n}$ are conditionally independent given $(\mathbf{X}_1,\cdots,\mathbf{X}_n)$ with conditional distribution function $G(.|\mathrm{X}_{R_1}),\cdots,G(.|\mathrm{X}_{R_n})$ respectively.

As remarked by Bhattacharya (1984), although his Lemma deals only with the induced order statistics, the key fact on which the proof of that Lemma depends is that the induced orders are a random permutation of $1,\cdots,n$ determined only by the $\mathbf{X}$'s. Therefore the extension above is trivial. Note that the ranking mechanism in RSS indeed produced a random permutation of the sampling units. Hence the above Lemma can be applied to the RSS sample. In particular, we have

$$Y_{(r)i} = \alpha + \boldsymbol{\beta}^T \mathbf{X}_{(r)i} + \varepsilon_{ri} \tag{3.3}$$

$$r = 1,\cdots,n, \quad i = 1,\cdots,m$$

where ε_{ri} are independent identically distributed as the ε in (3.2) and are independent of $\mathbf{X}_{(r)i}$. Thus we can estimate α and $\boldsymbol{\beta}$ by least square method based on (3.3)

$$\overline{\mathbf{X}}_{RSS} = \frac{1}{mn}\sum_{r=1}^{n}\sum_{i=1}^{n}\mathbf{X}_{(r)i},\ \overline{Y}_{RSS} = \frac{1}{mn}\sum_{r=1}^{n}\sum_{i=1}^{n}Y_{(r)i}$$

$$X_{RSS} = (\mathbf{X}_{(1)1},\cdots,\mathbf{X}_{(1)m},\cdots\mathbf{X}_{(n)1},\cdots\mathbf{X}_{(n)m})'$$

$$Y_{RSS} = (y_{(1)1},\cdots,y_{(1)m},\cdots y_{(n)1},\cdots y_{(n)m})'.$$

The least squares estimates of α and $\boldsymbol{\beta}$ based on (3.3) are then given by

$$\hat{\alpha}_{RSS} = \overline{\mathbf{Y}}_{RSS} - \hat{\boldsymbol{\beta}}'_{RSS}\,\overline{\mathbf{X}}_{RSS} \tag{3.4}$$

$$\hat{\boldsymbol{\beta}}_{RSS} = \frac{[\mathbf{X}'_{RSS}(\mathbf{I} - \frac{\mathbf{11}'}{mn})\mathbf{Y}_{RSS}]}{[\mathbf{X}'_{RSS}(\mathbf{I} - \frac{\mathbf{11}'}{mn})\mathbf{X}_{RSS}]} \tag{3.5}$$

It is obvious that $\hat{\alpha}_{RSS}$ and $\hat{\boldsymbol{\beta}}_{RSS}$ are unbiased. Since both $\hat{\alpha}_{RSS}$ and $\hat{\boldsymbol{\beta}}_{RSS}$ are of the form of smooth-function-of-means, as estimates of α

and $\boldsymbol{\beta}$, they are, at least asymptotically, as good as their counterparts based on SRS. More specific conclusion can be made from the explicit expressions of the variances of $\hat{\alpha}_{RSS}$ and $\hat{\boldsymbol{\beta}}_{RSS}$. These variances can be derived as

$$Var(\hat{\alpha}_{RSS}) = \sigma_\varepsilon^2 \; E\left[\frac{1}{mn} + \overline{\mathbf{X}}_{RSS}^T [\mathbf{X}_{RSS}' (\mathbf{I} - \frac{\mathbf{11'}}{mn}) \mathbf{X}_{RSS}]^{-1} \overline{\mathbf{X}}_{RSS}\right]$$

$$Var(\hat{\boldsymbol{\beta}}_{RSS}) = \sigma_\varepsilon^2 \; E\left[\, [\mathbf{X}_{RSS}' (\mathbf{I} - \frac{\mathbf{11'}}{mn}) \mathbf{X}_{RSS}]^{-1}\right]$$

where the expectations are taken with respect to the distribution of the X's.

3.2 Regression estimate of the mean of Y with RSS

Let

$$\overline{\mathbf{X}}_T = \frac{1}{mn^2} \sum_{r=1}^{n} \sum_{j=1}^{n} \sum_{i=1}^{m} \mathbf{X}_{rji}$$

We can define another estimate of μ_Y rather than $\overline{Y}_{RSS}$. The estimate is called the RSS regression estimate and is define as

$$\hat{\mu}_{RSS.REG} = \overline{\mathbf{Y}}_{RSS} - \hat{\boldsymbol{\beta}}'_{RSS} (\overline{\mathbf{X}}_T - \overline{\mathbf{X}}_{RSS}) \tag{3.6}$$

The RSS regression estimate of μ_Y is unbiased and its variance can be obtained as

$$Var(\hat{\mu}_{RSS.REG}) = \frac{\sigma_\varepsilon^2}{mn}\{1 + \Delta_{RSS}\} + \frac{1}{mn^2} \boldsymbol{\beta}' \Sigma \boldsymbol{\beta} \tag{3.7}$$

where

$$\Delta_{RSS} = E[mn(\overline{\mathbf{X}}_T - \overline{\mathbf{X}}_{RSS})'[(\overline{\mathbf{X}}_{RSS}(\mathbf{I} - \frac{\mathbf{11}'}{mn})\mathbf{X}_{RSS}]^{-1}(\overline{\mathbf{X}}_T - \overline{\mathbf{X}}_{RSS})]$$

If the ranked set sample is replaced by a simple random sample, we get an SRS regression estimate of μ_Y. The variance of the SRS regression estimate is also of the form (3.7) but with Δ_{RSS} replaced by the corresponding quantity Δ_{SRS} defined on the simple random sample.

REFERENCES

Abu-Dayyeh W. and Muttlak H.A. (1996): Using ranked set sampling for hypothesis testing on the scale parameter of the exponential and uniform distributions. *Pak. J. Statist.*, **12**, 131-138.

Balakrishnan, N. (1988): Recurrence relations for order statistics from *n* independent and nonidendically distributed random variables. *Ann. Inst. Statist. Math.*, **40**, 273-277.

Balakrishnan, N. and Cohen, A.C. (1991): *Order statistics and inference*, Academic Press, Inc., Boston.

Balakrishnan, N. and Li, T. (2005): BLUEs of parameters of generalized geometric distribution using ordered ranked set sampling. *Comm. Statist. Simu. Comt.*, **34**, 253-266.

Balakrishnan, N., Chan, P.S. and Varadan, J. (1992): Order statistics from extreme-value distributions. I: Tables of means, variances and co-variances, *Comm. Statist. Simu.Comt.*, **21**, 1199-1217.

Barlow, R.E. and Proschan, F. (1981): *Statistical theory of reliability and life testing*. Silver Spring, Maryland.

Bhattacharya, P.K. (1974): Convergence of sample paths of normalized sums of induced order statistics. *Ann. Statist.*, **2**, 1034-1039.

Bhattacharya, P.K. (1984): Induced order statistics theory and applications, In Krishnainh, P. R. and Sen, P. K., editors, *Handbook of Statistics,* **4**, 383-403. Elsevier Science Publishers.

Bhoj, D.S. and Ahsanullah, M. (1996): Estimation of parameters of generalized geometric distribution using ranked set sampling. *Biometrics,* **52**, 685-694.

Bickel, P.J. (1967): Some contribution to the theory of order statistics. *Proc. IV- th Berkeley Symp.*, **1**, 57 – 91.

Chen, Z. (2000): The efficiency of ranked set sampling relative to simple random sampling under multi-parameter families. *Statistica Sinica,* **10**, 247-263.

Chen, Z., Bai, Z. and Sinha, B.K. (2004): *Ranked Set Sampling; Theory and Application*. Springer-Verlag, New York

Chuiv, N.N. and Sinha, B.K. (1998): On some aspects of ranked set sampling in parametric estimation. *Book of Statistics,* 17, 337-377, N. Balakrishnan and R. Rao, Eds., North Holland, New York.

David, H.A. (1973): Concomitants of order statistics. *Bull. Inst. Internal. Statist.,* **45,** 295-300.

David, H.A. (1981): *Order Statistics.* John Wiley, New York.

David, H.A. and Groenevld, R.A. (1982): Measures of local variation in a distribution: Expected length of spacing of order statistics. *Biometrics*, **69**, 227-232.

David, H.A. and Levine, D.N. (1972): Ranked set sampling in the presence of judgement error. *Biometrics,* **28**, 545-553.

David, H.A., O'Connell, M.J. and Yang, S.S. (1977): Distribution and expected value of the rank of a concomitant of an order statistic. *Ann. Statist.*, **5**, 216-223.

Dell, T.R. (1969): The theory of some applications of ranked set sampling, *Ph. D. Thesis,* University of Georgia, Athens, GA.

Dell, T.R. and Clutter, J.L. (1972): Ranked set sampling theory with order statistics background. *Biometrics,* **28**, 545—555.

El-Neweihi, E. and Sinha, B.K. (2000): Reliability estimation based on ranked set sampling. *Comm. Statist. Theory-Methods,* **29 (7),** 1583-1595.

Fie, H., Sinha, B.K. and Wu, Z. (1994): Estimation of parameters in two parameter Weibull and extreme-value distributions using raked set sampling. *J. Statist. Res.,* **28**, 149-161.

Ghitany, M.E. (2005): On reliability estimation based on ranked set sampling. *Comm. Statist. Theory-Methods,* **34**, 1213-1216.

Ghosh, M. and Sen, P.K. (1971): On a class of rank order tests for regression with partially informed stochastic predictors. *Ann. Math. Statist.*, **42**, 650-661.

Hossain, S.S. and Muttlak, H.A. (2000): MVLUE of population parameters based on ranked set sampling, *App. Math. Com.*, **108**, 167-176.

Kaur, A., Patil, G.P., Sinha: A.K. and Taillie, C. (1999). Ranked set sampling: a bibliography. *Environ. Ecol. Stat.,* **6**, 91-98.

Lam, K., Sinha, B.K. and Wu, Z. (1995): Estimation of location and scale parameters of a logistic distribution using a ranked set sampling, *Papers in Honor of* Herbert A. David (edited by Nagaraja, Sen and Morrison), 187-197.

Lloyd, E.H. (1952): Least square estimation of location parameters using order statistics. *Biometrika*, **39**, 88-95.

Mann, N.R., Schafer, R.E. and Singpurwalla, N.D. (1974): *Methods for statistical analysis of reliability and life data*, John Wiley & Sons, New York.

Marshall, A. W. and Olkin, I. (1979): Inequalities, *Theory of majorization and its application,* Acadimic Press, New York.

McIntyre, G.A. (1952): A method of unbiased selective sampling, using ranked sets. *Austr. J. Agricultural Res.,* **3**, 385-390.

Mizuno, H. (1974): 'Mathematical Method in Sampling' *Symposium held at Chiba University,* Japan

Muttlak, H.A. (2003): Modified ranked set sampling methods. *Pak. J. Statist.,* **19(3)**, 315-323.

Muttlak, H.A. and McDonald, L.L. (1990a): Ranked set sampling with size biased probability of selection. *Biometrics,* **46**, 435-445.

Muttlak, H.A. and McDonald, L.L. (1990b): Ranked set sampling with respect to a concomitant variables and with size biased probability of selection, reliability estimation based on ranked set sampling. *Comm. Statist. Theory-Methods,* **19 (1)**, 205-219.

Muttlak, H.A. and McDonald, L.L. (1992). Ranked set sampling and the line-intercept method: a more efficient procedure. *Biom. J.*, **3**, 329-346.

Patil, G.P., Sinha A.K. and Taillie C. (1994): Ranked Set Sampling, In: G.P. Patil and C.R. Rao, eds. *Handbook of Statistics*, **12**, 167-200.

Patil, G.P., Sinha, A.K. and Taille, C. (1993): Relative precision of Ranked set sampling: a comparison with the regression estimator. *Environmetrics*, **4(4)**, 399-412.

Ridout, M.S. and Cobby, J.M. (1987): Ranked set sampling with non-random selection of sets and errors in ranking. *App. Statist.,* **36**, 145-152.

Sarndal, C.E., Swensson, B.I. and Wretman, J. (1992): *Model Assisted Survey Sampling*, Springer-Verlag, 652—659.

Sen, P.K. (1976): A note on invariance principles for induced order statistics. *Ann. Probab.*, **4**, 474-479.

Sen, P.K. (1981): some invariance principles for mixed rank statistics and induced order statistics and some applications. *Commun. Statist. Theory-Methods*, **10**, 1691-1718.

Shen W.H. (1994): Use of ranked set sampling for test of a normal mean. *Calcutta Statist. Assoc. Bull.*, **44**, 183-193.

Silva, P.L.D.N. and Skinner, C.J. (1997): Variable selection for regression estimation in finite populations. *Survey Methodology,* **23**, 23-32.

Sinha, B.K., Lam, K. and Wu, Z. (1995): Estimation of location and scale parameters of logistic distribution using a ranked set sampling. *In collected Essays in Honor of* David, H.A., Nagaraja, H.N., Sen, P.K. and Morrison, D.F. editors, 187-197.

Sinha, Bimal K. Sinha Bikas k. and Purkayastha, S. (1996): On some aspects of ranked set sampling for estimation of Normal and Exponential parameters. *Statistical Decisions,* **14**, 223-240.

Stokes, S.L. (1976): An investigation of the consequences of ranked set sampling, *Ph.D. Thesis*, University of North Carolina, Chapel Hill, NC.

Stokes, S.L. (1977): Ranked set sampling with concomitant variables. *Comm. Statist. Theory-Methods*, **6**, 1207-1211.

Stokes, S.L. (1980): Estimation of variance using judgment ordered ranked set samples. *Biometrics,* **36**, 35-42.

Stokes, S.L. (1986): Ranked set sampling. In: S. Kotz et al., eds. *Encyclopaedia of Statistical Sciences*, Wiley, New York, 585-588.

Stokes, S.L. And Sager, T. W. (1988): Characterization of a ranked set sample with application to estimating distribution functions. *J. Amer. Statist. Assoc.*, **83**, 374-381.

Takahasi, K. (1969): On the estimation of the population mean based on ordered sample from an equicorrelated multivariate distribution. *Ann. Inst. Statist. Math.*, **21**, 249-255.

Takahasi, K. And Wakimoto, K. (1968): On unbiased estimates of the population mean based on the sample stratified means of ordering. *Ann. Inst. Statist. Math.*, **20**, 1-31.

Tietjen, G.L. Kahaner, D.K. and Beckman, R.J. (1977): Variances and co-variances of the normal order statistics for sample size 2 to 50. *Selected Table in Mathematical Statistics,* **5**, 1-73.

Tukey, J.W. (1958): A problem of Berkson and minimum variance orderly estimators. *Ann. Math. Statist.*, **29**, 588-592.

Vaughan, R.J. and Venables, W.N. (1972): Permanent expression for order statistics densities. *J. Roy. Statist. Soc. Ser.*, **B 34**, 308-310.

Yanagawa, T. and Chen, S.H. (1980): The MG-procedure in ranked set sampling. *J. Statist. Plann. Infer.*, **4**, 33-44.

Yanagawa, T. and Shirahata, S. (1976): Ranked set sampling theory with selective probability matrix. *Aust. J. Statist.*, **18**, 45-52.

Yang, S.S. (1977): General distribution theory of concomitant of order statistics. *Ann. Statist.*, **5**, 996-1002.

Yu. P.L.H. and Lam, K. (1997): Regression estimator in ranked set sampling. *Biometrics*, **53**, 1070-1080.